

多言語検索ツール *AntPConc* の機能と使い方

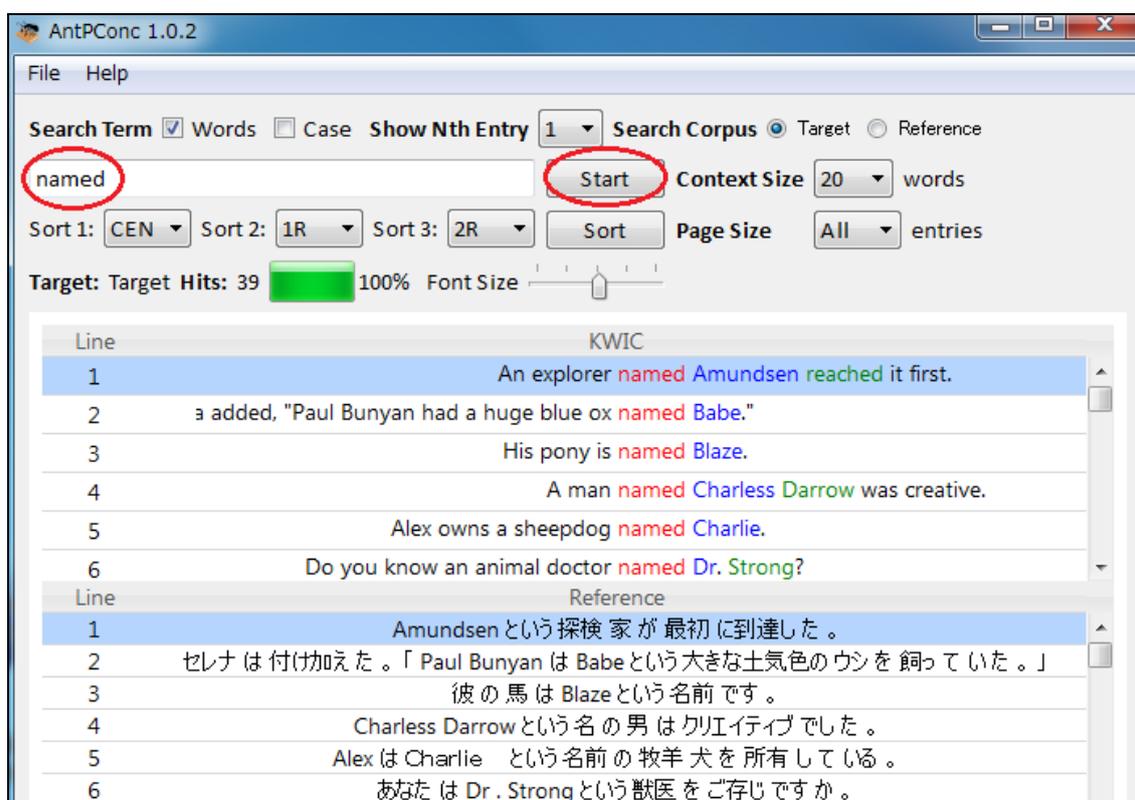
(Ver. 1)

○ *AntPConc* とは

AntPConc は、コンコーダンスーという種類の検索ツールであり、検索対象である多言語コーパスをデータベースに読み込み、検索を行う。本マニュアルでは、英語と日本語の 2 言語のコーパスを読み込んで利用する場合について述べる。

AntPConc で検索すると、指定した検索語を中心に据えた Key Word In Context (KWIC) と呼ばれる表示形式のコンコーダンス画面 (コンコーダンスライン) が、英語文と日本語文の二言語について同時に得られる。

検索したい語句、たとえば “named” を Start ボタンの左側のボックスに入れて、Start ボタンをクリックするか ENTER キーを押すと、このような検索結果が得られる。画面上半分の KWIC 画面に “named” を含む英語文が表示され、それらの英文に対応する日本語文が、画面下半分の Reference 画面に表示される。

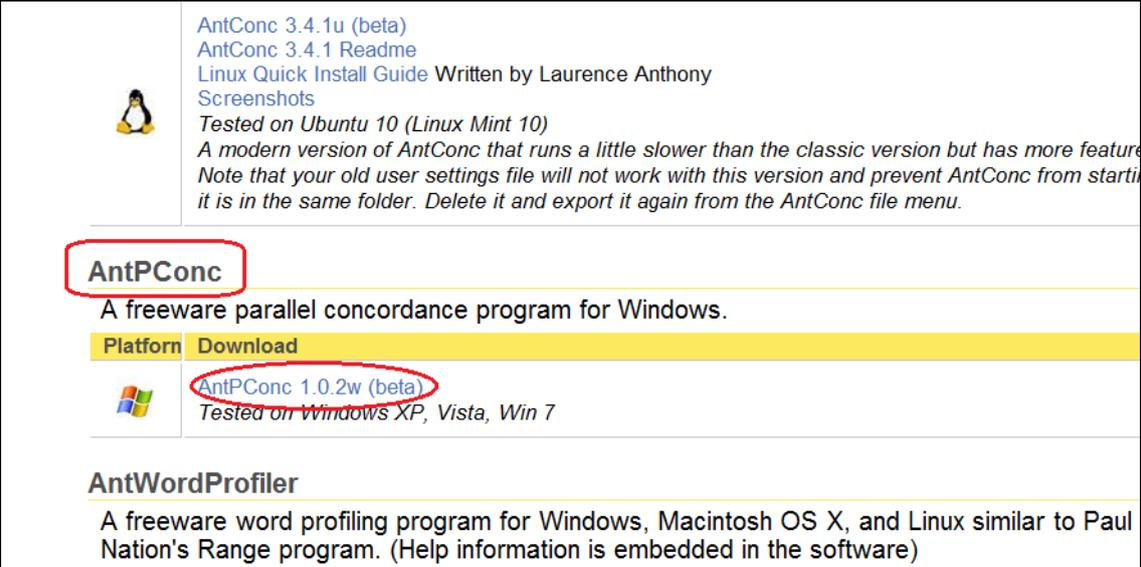


本マニュアルでは、*AntPConc* に実装された機能とその使い方を説明する。さらに、*AntPConc* で検索に用いる自作の多言語パラレルコーパスの作成方法を紹介する。

○ 準備編

AntPConc のダウンロード

http://www.antlab.sci.waseda.ac.jp/software.html#antpconc にアクセスするか、あるいはインターネットで「Anthony Laurence」を検索して最初に得られる検索結果をダブルクリックし、出てきた右上の Latest Software Releases の中から、AntPConc の最新版をみつけてダブルクリックする。次に、AntPConc の Download の文字の下のブルーの文字の部分 (AntPConc 1.0.2w (beta)) を右クリックする。「対象をファイルに保存」を選び、保存先を「デスクトップ」に選び「保存」を押す。



	AntConc 3.4.1u (beta) AntConc 3.4.1 Readme Linux Quick Install Guide Written by Laurence Anthony Screenshots <i>Tested on Ubuntu 10 (Linux Mint 10)</i> <i>A modern version of AntConc that runs a little slower than the classic version but has more features.</i> <i>Note that your old user settings file will not work with this version and prevent AntConc from starting if it is in the same folder. Delete it and export it again from the AntConc file menu.</i>
AntPConc A freeware parallel concordance program for Windows.	
Platform	Download
	AntPConc 1.0.2w (beta) <i>Tested on Windows XP, Vista, Win 7</i>
AntWordProfiler A freeware word profiling program for Windows, Macintosh OS X, and Linux similar to Paul Nation's Range program. (Help information is embedded in the software)	

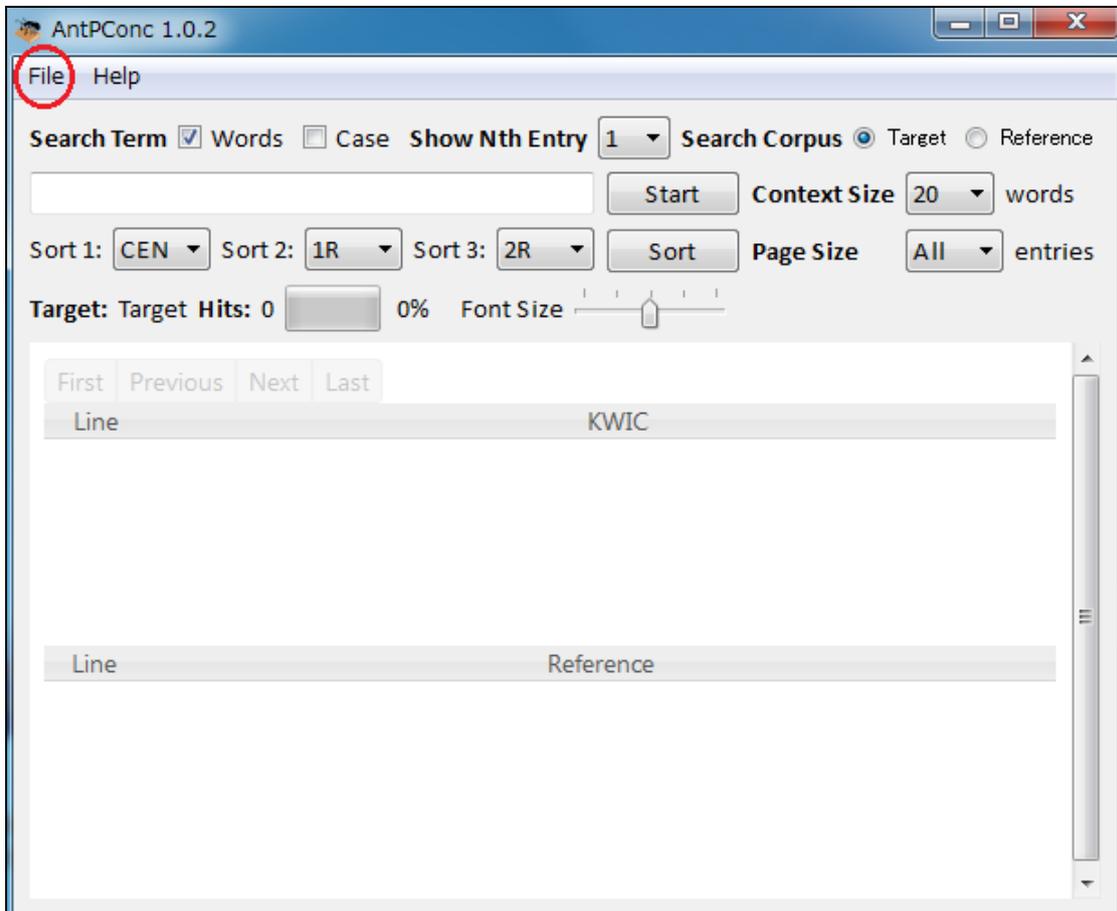
検索ファイル（コーパスデータ）の読み込みと保存：

AntPConc を使って検索するには、コーパスデータを用意する必要がある。英語コーパス（ここでは 607-en.txt と呼ぶ）と日本語コーパス（607-ja.txt と呼ぶ）をデスクトップにおく。AntPConc もデスクトップにおく。

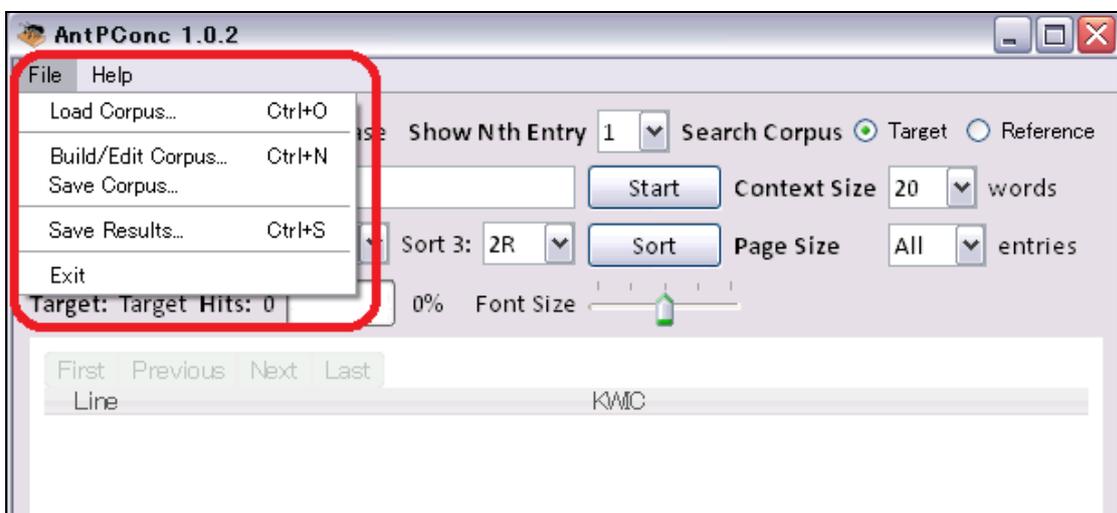
あるいは、USB ドライブに AntPConc、英語コーパス、日本語コーパスを保存しておいて使用することも可能である。ただし、上記のファイルを USB ドライブに直接入れること、すなわちフォルダに入れないよう留意する。

① File

AntPConc のアイコンをダブルクリックすると、初期画面が得られる。

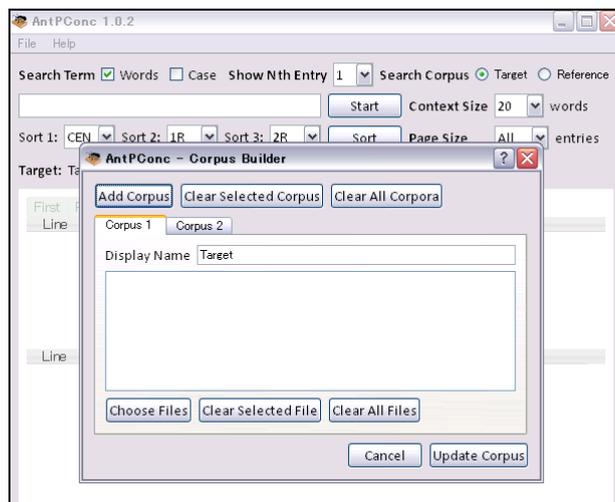


続いて、File ボタンをクリックすると、メニュー画面が現れる。

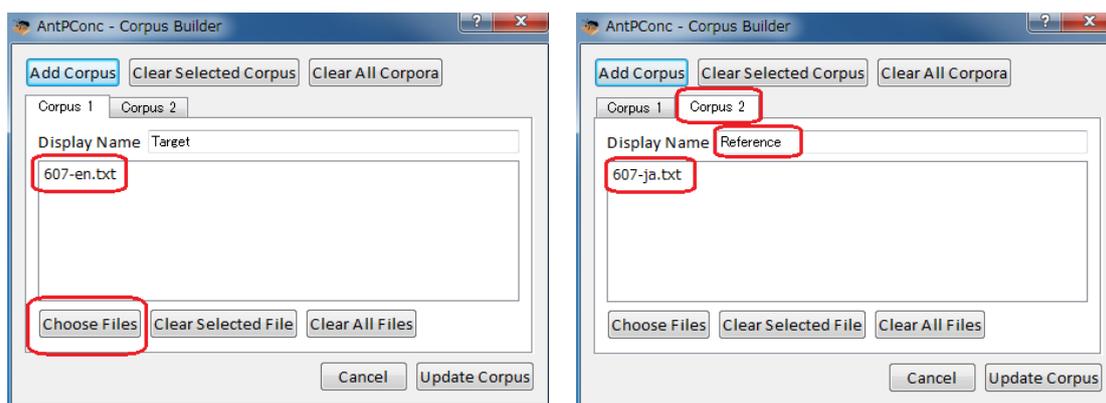


② Build/Edit Corpus

メニュー画面の2番目のメニューBuild/Edit Corpusにカーソルを合わせてクリックする。
'AntPConc Corpus Builder'でChoose Filesをクリックする。



デスクトップから「607-en.txt」（英語コーパスデータ）を指定して、「開く」をクリックする。ウィンドウに「607-en.txt」が入る。続いて、Corpus 2をクリックし、Choose Files をクリックして、デスクトップから「607-ja.txt」（日本語コーパスデータ）をクリックする。ウィンドウに「607-ja.txt」が入る。



なお、Choose Filesの作業を早めるため、「ドラッグアンドドロップ」を使うことも可能である。また、各コーパスをAntPConcの画面で区別するため、Display Nameで表示するコーパス名を入力することができる。ここでUpdate Corpusのボタンをクリックすれば、検索を開始することができる

③ Save Corpus

複数のコーパスデータを読み込む手間を次回から省くために、読み込んだ複数コーパスデータを1つの「データベースファイル」として保存しておく便利な方法について述べる。

②の手順でコーパスデータを読み込んだ後、File の3番目のメニューの Save Corpus にカーソルを合わせてクリックする。データベースファイルの保存場所と保存ファイル名（ピリオドで区切られた右側の文字列部分の拡張子に.db が付く）を指定して「保存」を押す。ここでは、「607_corpus.db」という名前で保存した。

④ Load Corpus

前述の③Save Corpus の手順を済ませておくと、次回からまず、Load Corpus をクリックして、ファイル名(本稿では「607_corpus.db」)を指定し、「開く」を選択するとすぐに *AntPConc* が検索可能な状態になる。

授業や自習で *AntPConc* を使用する時には、データベースファイル（「607_corpus.db」）を学習者に配布するとスムーズに検索作業に入ることができる。

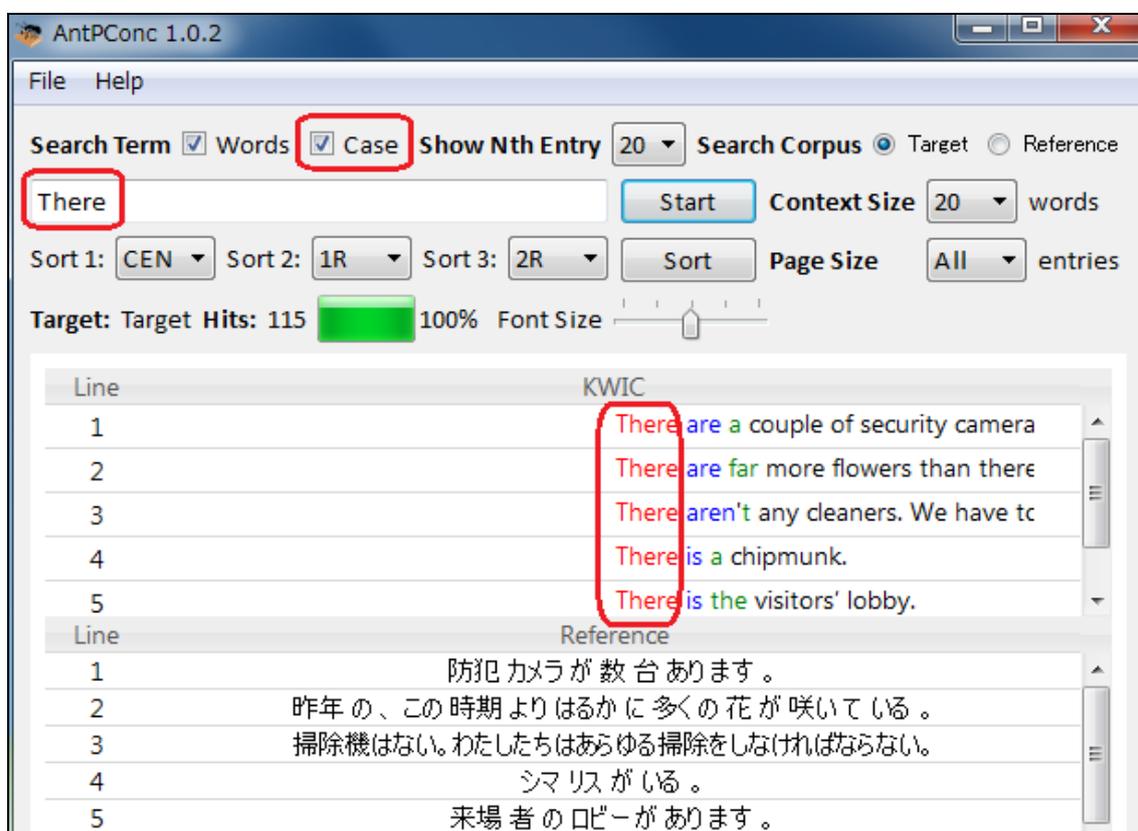
○ 検索編

AntPConc のデフォルト設定で十分な検索が可能であるが、さらに詳細な検索設定が可能である。

Search Term : 検索語の大文字・小文字の区別

AntPConc のデフォルト設定では、検索の際、英語の大文字と小文字は区別しない。小文字で検索語句を “must” と入力しても、大文字の “Must” も検索結果として出力される。

一方、Search Term の Case の左の をクリックして Case に設定すると、大文字と小文字を区別した検索が実行される。例えば、 Case 設定で “There” を検索すると、小文字の “there” を含む文は出力されず、“There is/are” の文を出力することが可能である。

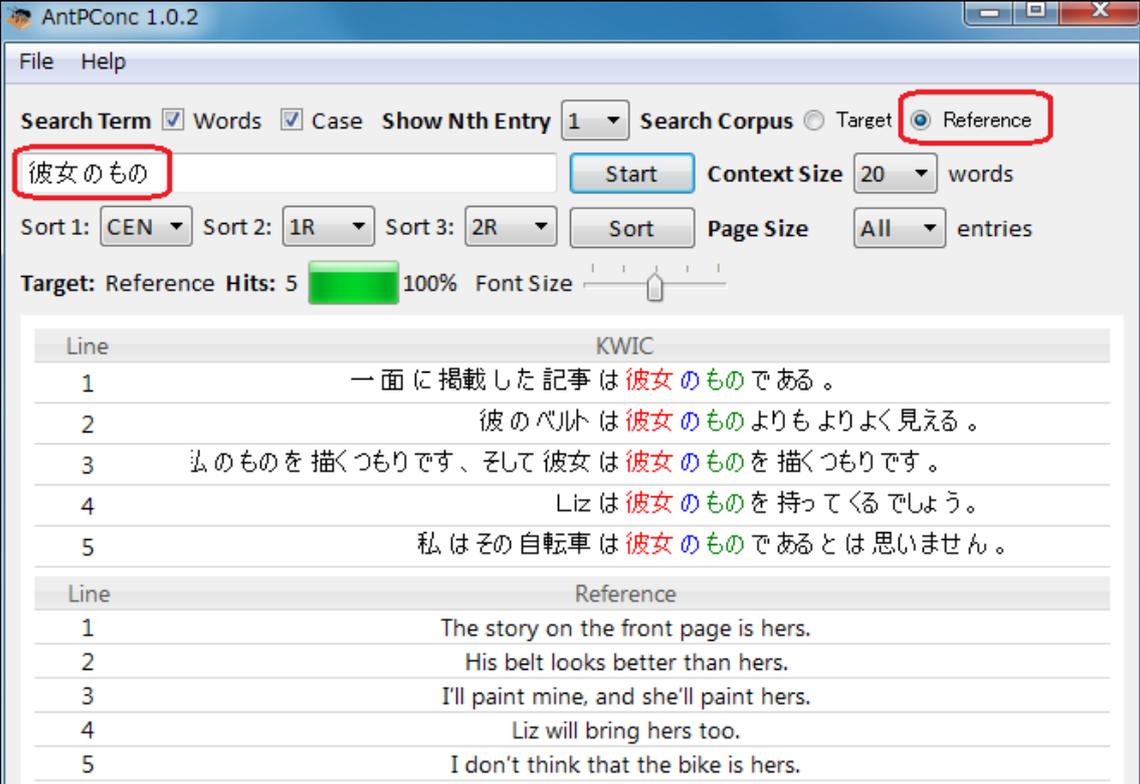


Search Corpus : 検索コーパスの選択

AntPConc を用いると、複数の言語のコーパスを検索することができる。Search Corpus のラジオボタンを指定のコーパス（ここでは Target あるいは Reference）に設定することによって、コンコーダンス画面に表示する言語を選択できる。3 つ以上のコーパスを取り込んでいる場合、各コーパスの指定名がここに表示される。

AntPConc を使用する多くの場合、Search Corpus は Target（英語コーパス）が選択されており、検索結果が KWIC 画面（上半分）に表示される。また、Reference（日本語コーパス）の検索結果が Reference 画面（下半分）に表示される。

一方、Search Corpus を日本語コーパスに設定することも可能である。検索語句として“彼女のもの”を検索した結果を示した。日本語の“彼女のもの”を含む検索結果が KWIC 画面に、“hers”を含む英語の検索結果が Reference 画面に表示される。



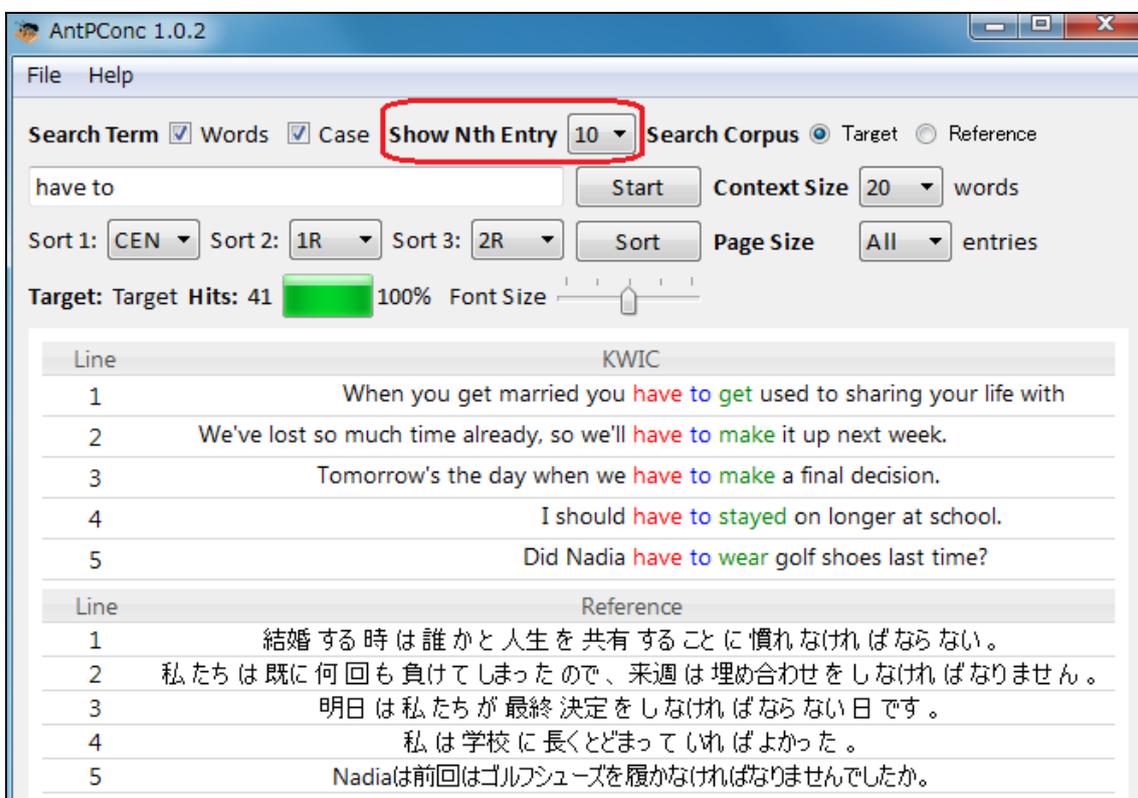
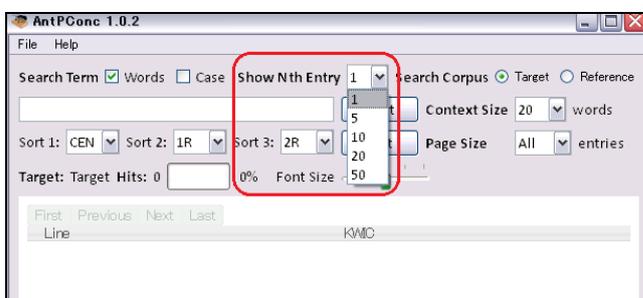
The screenshot shows the AntPConc 1.0.2 interface. The search term "彼女のもの" is entered in the Search Term field. The Search Corpus is set to Reference. The search results are displayed in two sections: KWIC and Reference.

Line	KWIC
1	一面に掲載した記事は彼女のものである。
2	彼のベルトは彼女のものよりもよりよく見える。
3	私のものを描くつもりです、そして彼女は彼女のものを描くつもりです。
4	Lizは彼女のものを持ってくるでしょう。
5	私はその自転車は彼女のものであるとは思いません。

Line	Reference
1	The story on the front page is hers.
2	His belt looks better than hers.
3	I'll paint mine, and she'll paint hers.
4	Liz will bring hers too.
5	I don't think that the bike is hers.

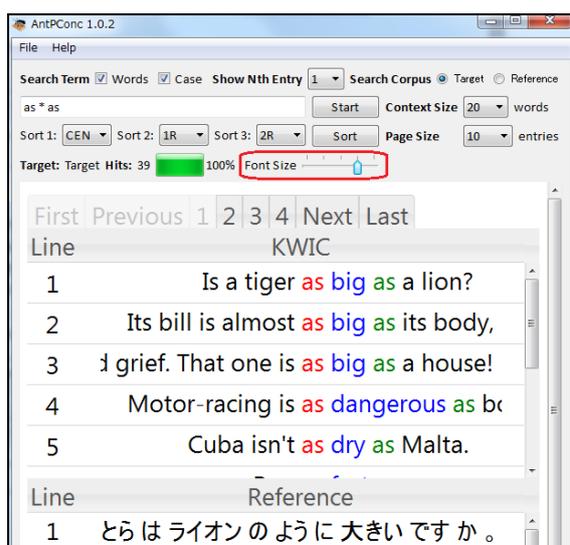
Show Nth Entry : 検索結果の表示件数

AntPConc のデフォルト設定では、Show Nth Entry は 1 に設定されており、すべての検索結果が表示される。検索結果の件数が多い時には、スクロールしてすべてのコンコーダンスラインを観察して、そこからルールを導き出すのは学習負担が大きい。そこで、Show Nth Entry を 5, 10, 20, 50 の中から選択して、検索結果を 5 文ごと、10 文ごと、20 文ごと、50 文ごとに表示することによって、コンコーダンスラインの概要を把握し易くすることが可能となる。例えば、“have to”の用法を観察する場合、Show Nth Entry を 10 に設定すると、合計 41 行の検索結果の表示件数が 5 行の例文表示となり、少数の英語と日本語文を対比した比較・観察が可能となる。



Font Size : フォントサイズの調節

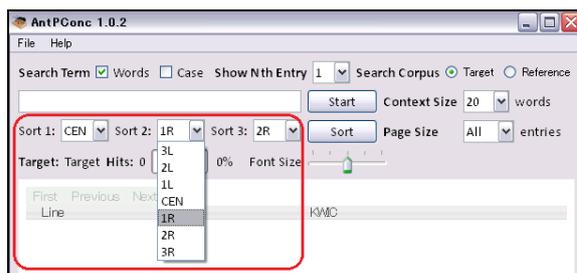
AntPConc では、Font Size の目盛をスライドさせることによってフォントサイズを調節することができる。フォントサイズを一目盛大きくすると、コンコーダンスラインが非常に見やすくなり、学習者はコンピュータ画面から顔を離すことが可能となる。また、机間巡視をする教師にも学習者のコンピュータ画面が見やすくなり、学習活動の進捗状況の把握が容易になる。



Sort : 並べ替え

AntPConc では、検索結果を観察しやすくするために、検索語および検索語の左右の語をABC順に並べ替えるソート (Sort) 機能が付いている。検索結果の表示順序を、検索語から左右3番目の語まで、第1 (Sort 1), 第2 (Sort 2), 第3 (Sort 3) の3つのソートキーごとに優先順位を指定することができる。

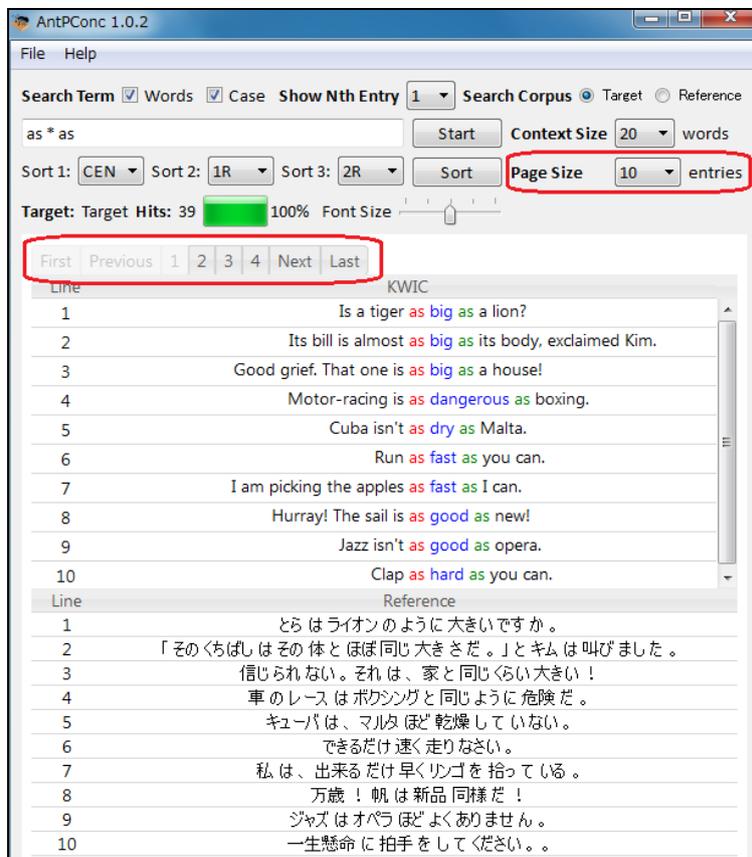
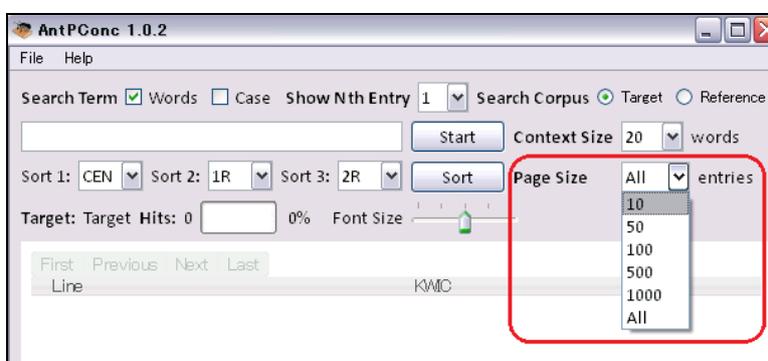
デフォルトでは、Sort 1 が CEN (検索語), Sort 2 が 1R (検索語の右1語目), Sort 3 が 2R (検索語の右2語目) (いわゆる右ソート) と設定されており、Search ボタンをクリックするとその優先順位でソートした結果が表示される。また、一度出力されたコンコーダンスラインに対して、新たにソートの基準を変えて Search ボタンをクリックすると新たな表示順序に並べ替えることができる。



Page Size : ページサイズの調節

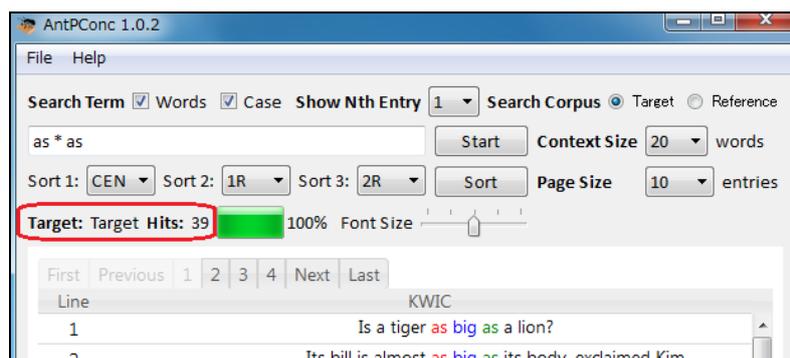
大量の出力結果を表示する際、1 ページ (1 画面) に表示する行数を選択することができる。AntPConc では、検索結果の Page Size を 10 行、50 行、100 行、500 行、1000 行、All の 6 通りから選択できる。デフォルトは All に設定されており、1 ページに検索結果のすべての行が表示されるので、スクロールして見る必要がある。

例えば、“as * as” の検索では、Page Size を 10 に設定した検索結果を示した。1 ページに 10 行ずつ結果が表示される。2 ページ目を見るには KWIC 画面の上に表示されたページボタンのタブの 2 をクリックする。3 をクリックすると 3 ページ目が表示される。スクロールしなくてよいので、出力結果の件数が多い時には、この機能が便利である。



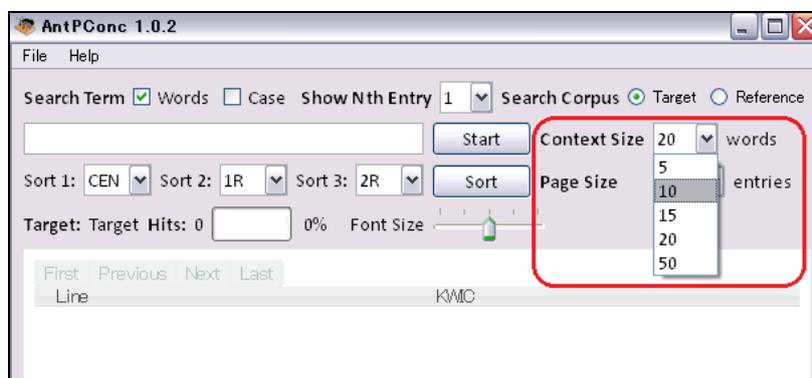
Hits : 検索ヒット数

Target Hits の右側に、検索結果の件数が表示される。例えば、教師のデモンストレーションの検索結果の件数を見ることで、学習者は自分が正しく検索できているかを自己チェックできる。また、検索結果の件数に応じて、Show Nth Entry 等の設定を変えて、コンコーダンスラインを観察することが可能となる。



Context Size : 1 行の長さの調節

Context Size は KWIC 画面での検索語句を含む左右の語数を示す。デフォルトでは 20 に設定されているが、5 words, 10 words, 15 words, 20 words, 50 words の中から選択できる。



その他

ワイルドカード : 検索語句の入力にはワイルドカード (wildcard) と呼ばれる特殊記号を利用することができる。「*」は任意の文字列を表し、例えば、*as * as* と入力すると、*as big as*, *as dangerous as*, *as dry as* などが検索できる。

ハイライト : KWIC 画面あるいは Reference 画面のコンコーダンスラインをクリックすると、対応する Reference 画面あるいは KWIC 画面の両方のコンコーダンスラインがハイライト表示される。

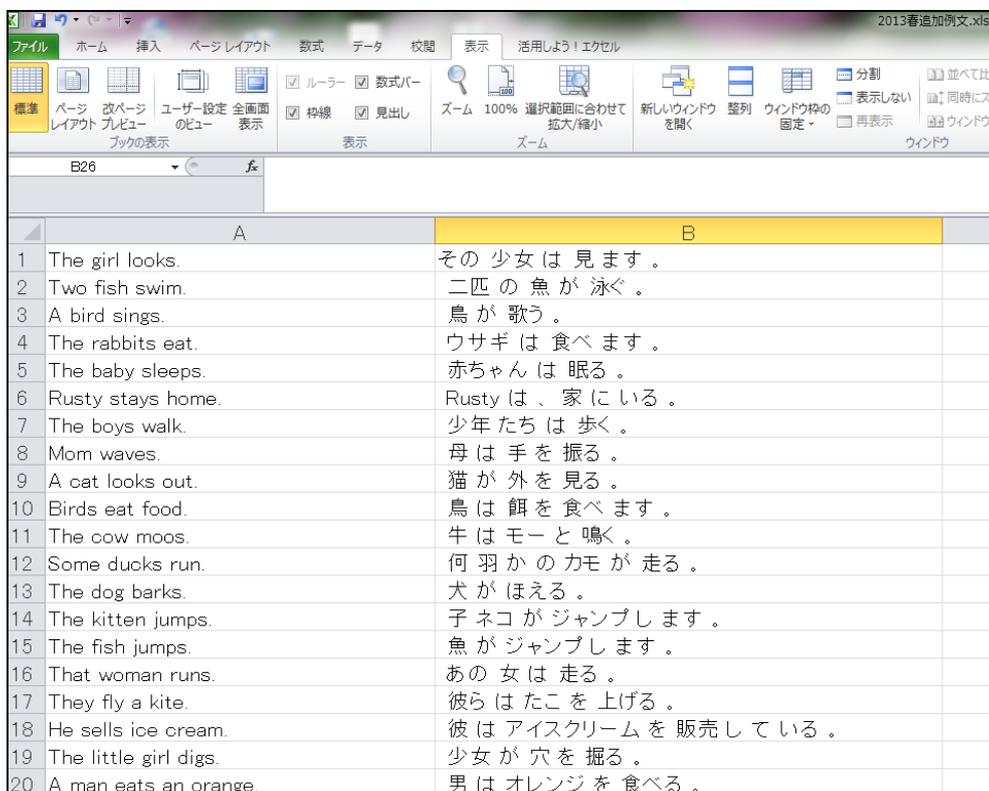
○ 自作コーパスデータの作成編

AntPConc を検索する英語と日本語の 2 言語コーパス（あるいは多言語コーパス）データの作成方法を述べる。自分で収集した英文と日本語文からコーパスデータが作成可能である。

Step 1: 「エクセル」で 1 つの列に英語の文をタイプする。隣の列に英語に対応する日本語の対訳文をタイプする。

Step 2: 英語列をコピーして、「メモ帳」に貼りつけ、UTF-8 形式で保存して、英語ファイルを作成する。次に、日本語列をコピーして、「メモ帳」に貼りつけ、UTF-8 形式で保存して、日本語ファイルを作成する。*AntPConc* で検索可能なコーパスデータが作成できる。

Step 3: このステップはとばすことも可能である。日本語テキストは、B列のように「分かち書き」をすると、検索の際、単語ごとにきれいに検索やソートが可能となる。分かち書きとは、文において、語の区切りに空白をはさんで区切って書き表すことである。英語の場合は単語と単語の間にスペースが入っているので分かち書きは不要である。分かち書きプログラムには、TECDICP (<http://www.te-com.biz/download/page1.html>) や Japanese Tutor (<http://www.vector.co.jp/soft/dl/winnt/edu/se462303.html>) が利用可能である。



	A	B
1	The girl looks.	その少女は見ます。
2	Two fish swim.	二匹の魚が泳ぐ。
3	A bird sings.	鳥が歌う。
4	The rabbits eat.	ウサギは食べます。
5	The baby sleeps.	赤ちゃんは眠る。
6	Rusty stays home.	Rustyは、家にいる。
7	The boys walk.	少年たちは歩く。
8	Mom waves.	母は手を振る。
9	A cat looks out.	猫が外を見る。
10	Birds eat food.	鳥は餌を食べます。
11	The cow moos.	牛はモーと鳴く。
12	Some ducks run.	何羽かのカモが走る。
13	The dog barks.	犬がほえる。
14	The kitten jumps.	子ネコがジャンプします。
15	The fish jumps.	魚がジャンプします。
16	That woman runs.	あの女は走る。
17	They fly a kite.	彼らはたこを上げる。
18	He sells ice cream.	彼はアイスクリームを販売している。
19	The little girl digs.	少女が穴を掘る。
20	A man eats an orange.	男はオレンジを食べる。

参考文献：

- 1) Anthony, L. 3AntPConc (Version 1.0.2) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp/>, 2013.
- 2) Anthony, L. Chujo, K., Kenji, Y. and Atsushi, M., Broadening the Scope of Parallel Corpus Tools: Using AntPConc in the DDL Classroom, Second Asia Pacific Corpus Linguistics Conference (APCLC 2014), 2014, March.
- 3) Chujo, K., Anthony, L. and Oghigian, K., Teaching Remedial Grammar through DDL using AntPConc, AACL 2013 - American Association for Corpus Linguistics, San Diego State University, California, USA, 2013, January.
- 4) Chujo, K., Chikako, N. and Oghigian, K., A DDL Approach to Remedial Grammar Using the AntPConc Parallel Concordancer, The 11th Asia TEFL International Conference, Ateneo de Manila University, Philippines, 2013, October.
- 5) 中條清美, 西垣知佳子, 「多言語検索ツール AntPConc を利用したリメディアル文法授業の実践」 外国語教育メディア学会 (LET) 第 53 回全国研究大会, 2013 年 8 月 9 日, 文京学院大学.
- 6) Chujo, K., Anthony, L., Oghigian, K. and Yokota, K., Teaching Remedial Grammar through Data-Driven Learning Using AntPConc, *Taiwan International ESP Journal*, 5(2), 2013, 65-90.

謝辞：本研究は平成 25-28 年度科学研究費補助金基盤研究（B）（25284108）を受けています。