

2009年6月27日

# AntConc Tutorial

## Quick Start and Reference

1. AntConc <sup>1</sup> 導入 :	pp. 1~4
2. 早速使ってみよう KWIC 機能 :	pp. 5~6
3. 出てきたものを並び替え SORT 機能 :	pp. 7~9
4. 検索に役立つ正規表現について :	p. 10
5. タグセット一覧とその活用法 :	pp. 11~12
6. 単語の分布を調べる Concordance Plot 機能 :	p. 13
7. まとまった語句を調べる Clusters 機能 :	pp. 14~15
8. 単語リストを作る Wordlist 機能 :	pp. 16~21
9. レンマ化(lemmatization)の方法 :	pp. 22~25
10. 語の共起関係を算出する Collocate 機能 :	pp. 27~29
11. 特徴語を抽出する Keyword List 機能 :	pp. 29~32

東京外国語大学大学院総合国際学研究所 言語文化専攻 博士後期課程1年

金田 拓

村上 明

---

<sup>1</sup> 2009年6月22現在、Version3.2.2をWindowsXPで動作確認

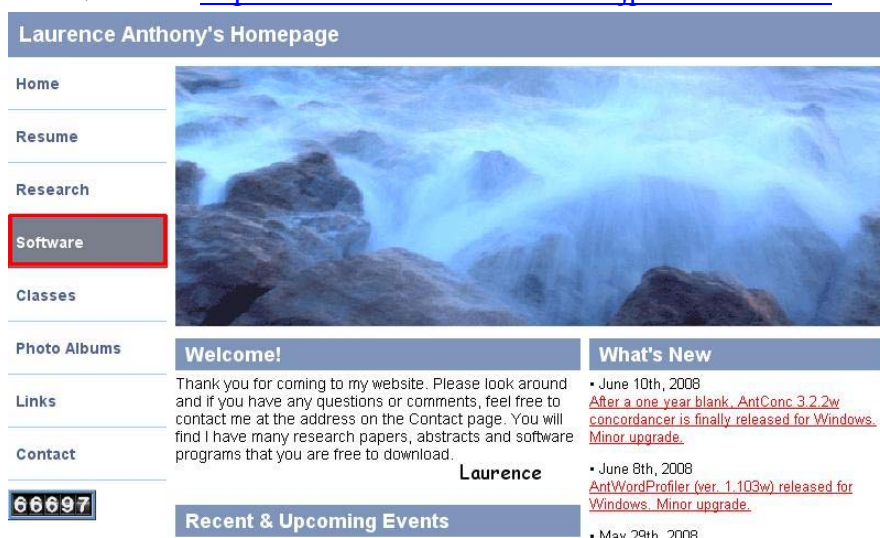
## 1. AntConc 導入

AntConc はフリーソフトでありながら高機能かつ使いやすいインターフェイスを備えた、優れたコンコーダンス・ソフトウェアです。Anthony LAURENCE 氏（早稲田大学）により開発されています。

Laurence 氏ホームページの「software」よりダウンロード可能です。

HP <http://www.antlab.sci.waseda.ac.jp/>

ソフト <http://www.antlab.sci.waseda.ac.jp/software.html>



Laurence Anthony's Homepage

Home

Resume

Research

**Software**

Classes

Photo Albums

Links

Contact

66697

Welcome!

Thank you for coming to my website. Please look around and if you have any questions or comments, feel free to contact me at the address on the Contact page. You will find I have many research papers, abstracts and software programs that you are free to download.

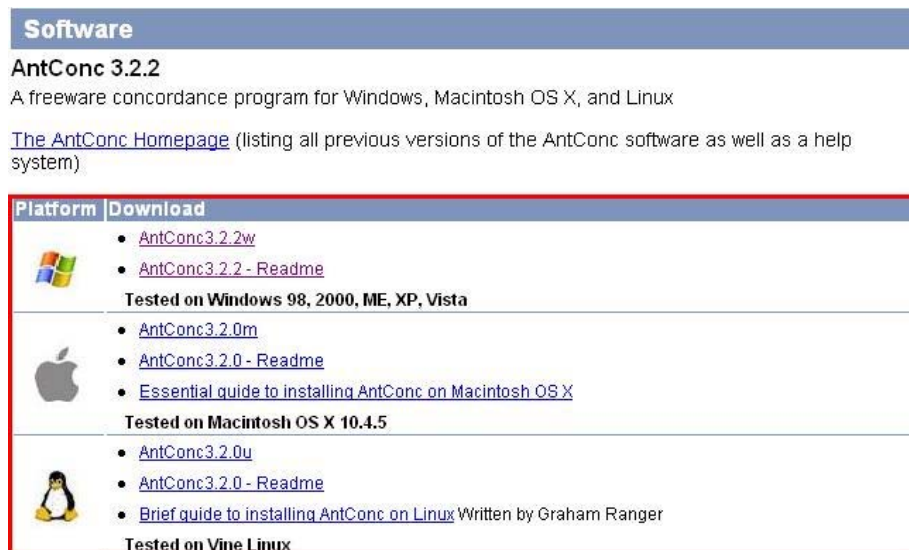
Laurence

What's New

- June 10th, 2008  
[After a one year blank, AntConc 3.2.2w concordancer is finally released for Windows. Minor upgrade.](#)
- June 8th, 2008  
[AntWordProfiler \(ver. 1.103w\) released for Windows. Minor upgrade.](#)
- May 29th, 2008

Recent & Upcoming Events

お使いの OS(Windows, Mac OS, Linux)に合わせたバージョンを選んでダウンロードします。






**Software**

**AntConc 3.2.2**

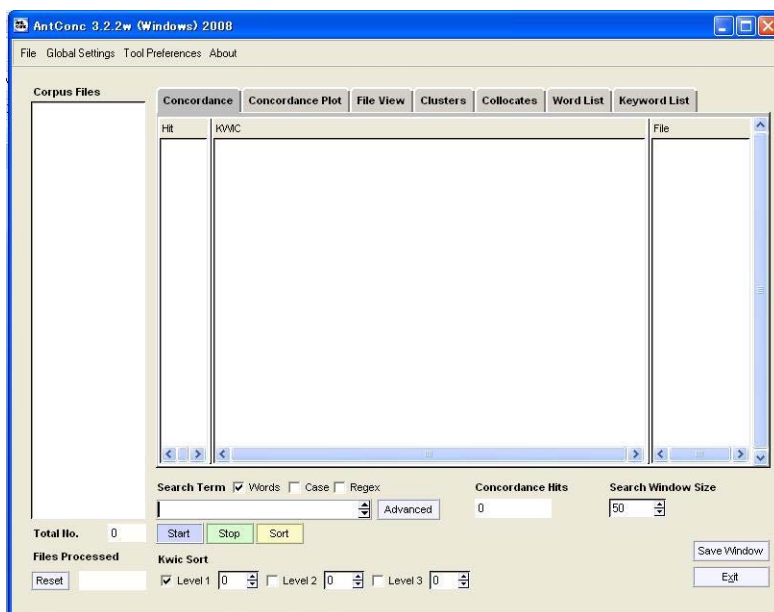
A freeware concordance program for Windows, Macintosh OS X, and Linux

[The AntConc Homepage](#) (listing all previous versions of the AntConc software as well as a help system)

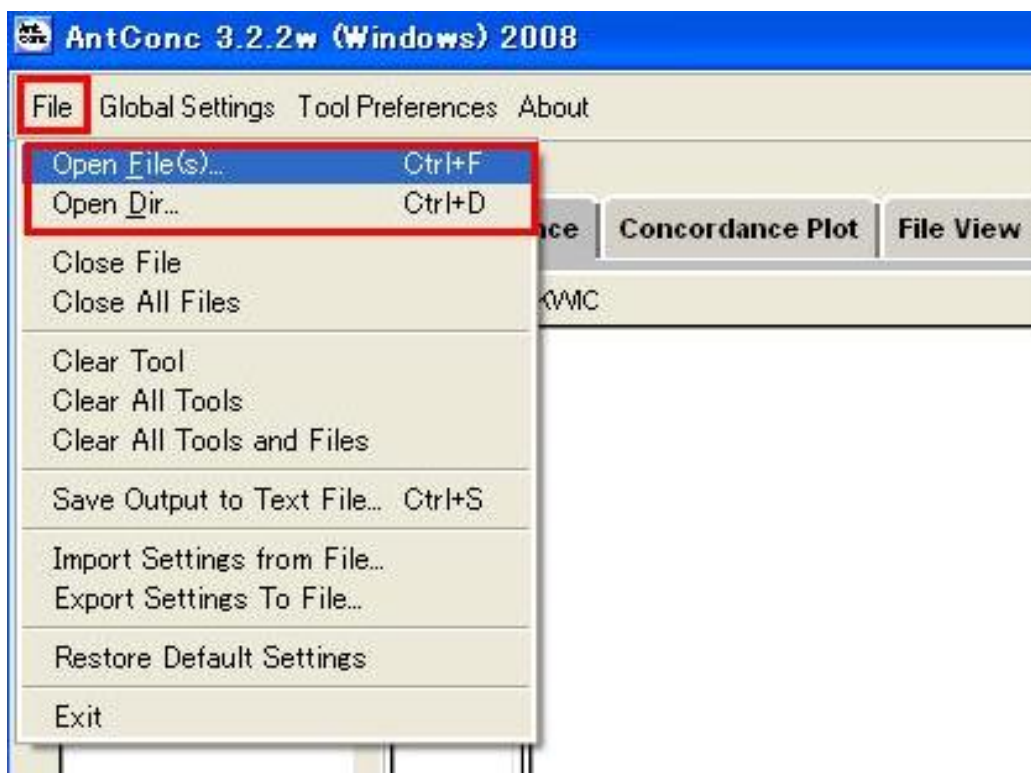
Platform	Download
	<ul style="list-style-type: none"><li>• <a href="#">AntConc3.2.2w</a></li><li>• <a href="#">AntConc3.2.2 - Readme</a></li></ul> <p><b>Tested on Windows 98, 2000, ME, XP, Vista</b></p>
	<ul style="list-style-type: none"><li>• <a href="#">AntConc3.2.0m</a></li><li>• <a href="#">AntConc3.2.0 - Readme</a></li><li>• <a href="#">Essential guide to installing AntConc on Macintosh OS X</a></li></ul> <p><b>Tested on Macintosh OS X 10.4.5</b></p>
	<ul style="list-style-type: none"><li>• <a href="#">AntConc3.2.0u</a></li><li>• <a href="#">AntConc3.2.0 - Readme</a></li><li>• <a href="#">Brief guide to installing AntConc on Linux</a> Written by Graham Ranger</li></ul> <p><b>Tested on Vine Linux</b></p>

ダウンロードしたファイルはお好きな場所に保存してください。

ダウンロードした実行ファイルをダブルクリックすると、下のような画面が出ます。

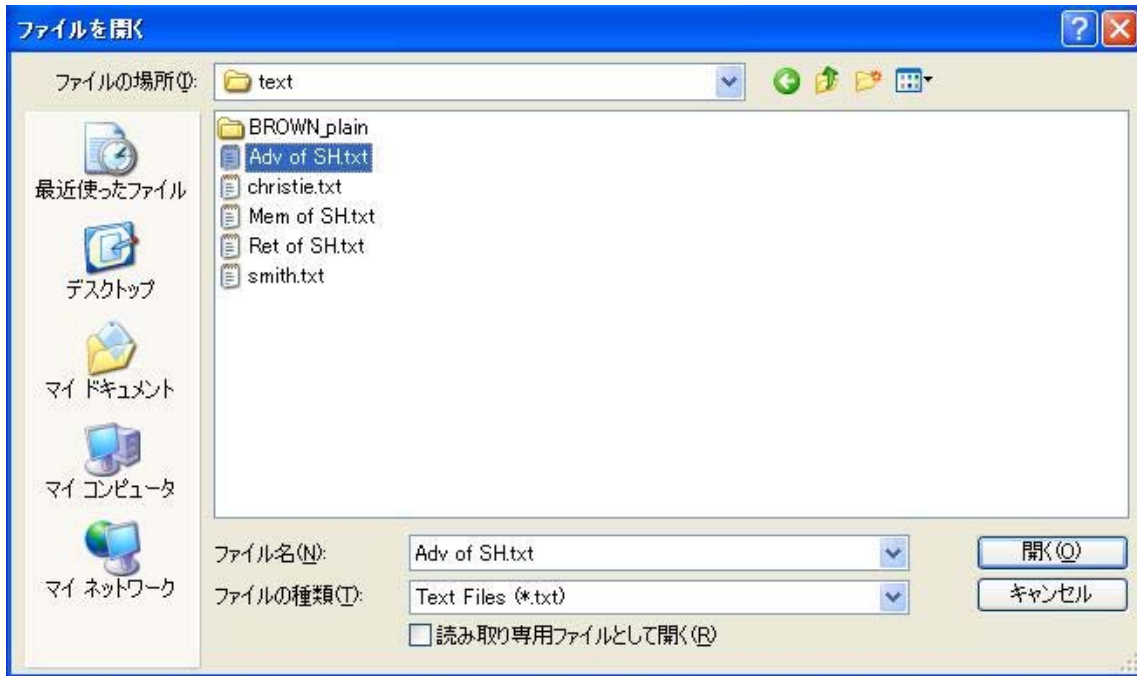


コンコーダンス・ソフトでは、分析するテキストを最初に指定します。

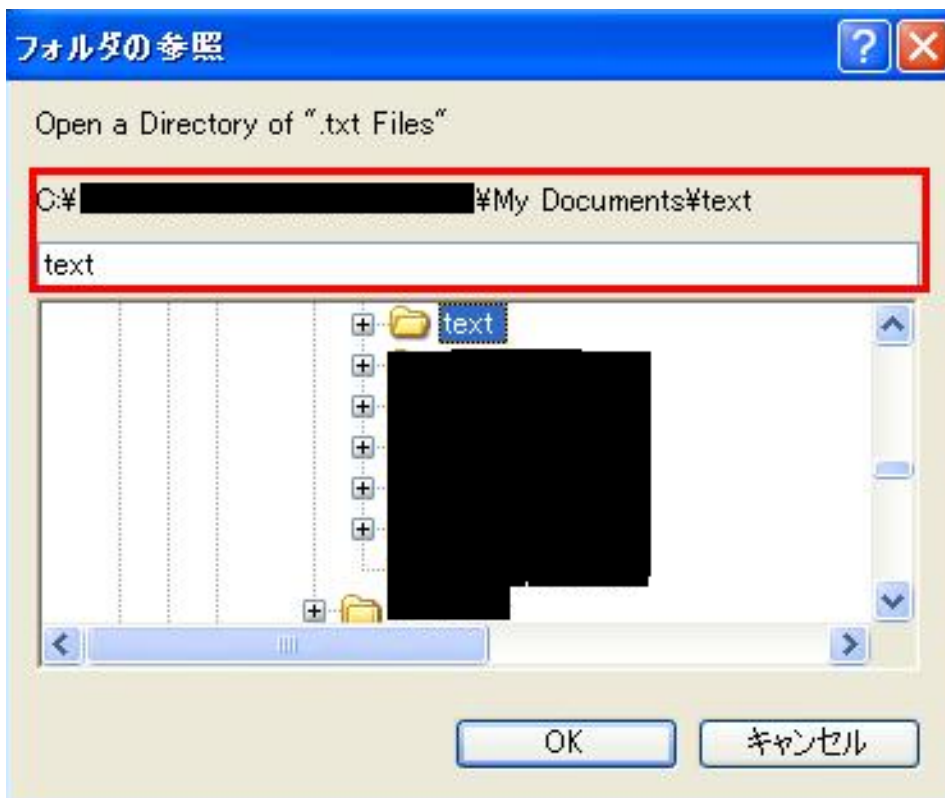


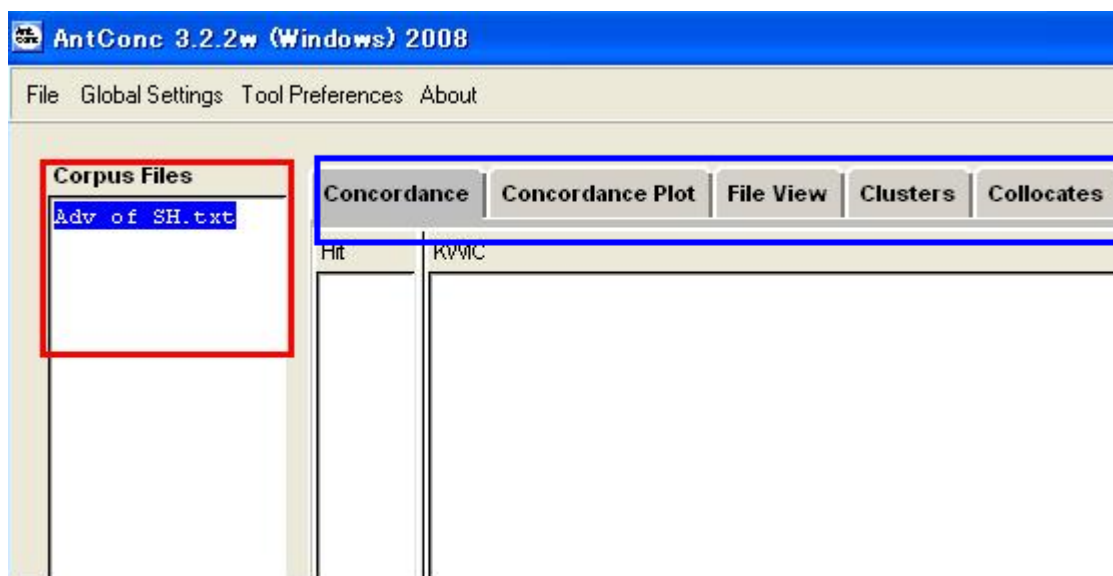
Open File(s)はテキストファイルを開くために、  
Open Dir はフォルダ単位で複数のファイルを開くために使います。

Open File(s)では以下のような選択画面になるので、分析したいファイルのある階層へ移動し、対象ファイルを選びます。(後でファイル追加も可能です)



Open Dir では、分析したいファイル群が含まれるフォルダを指定すると、その中のファイル全てが分析対象として選択されます。





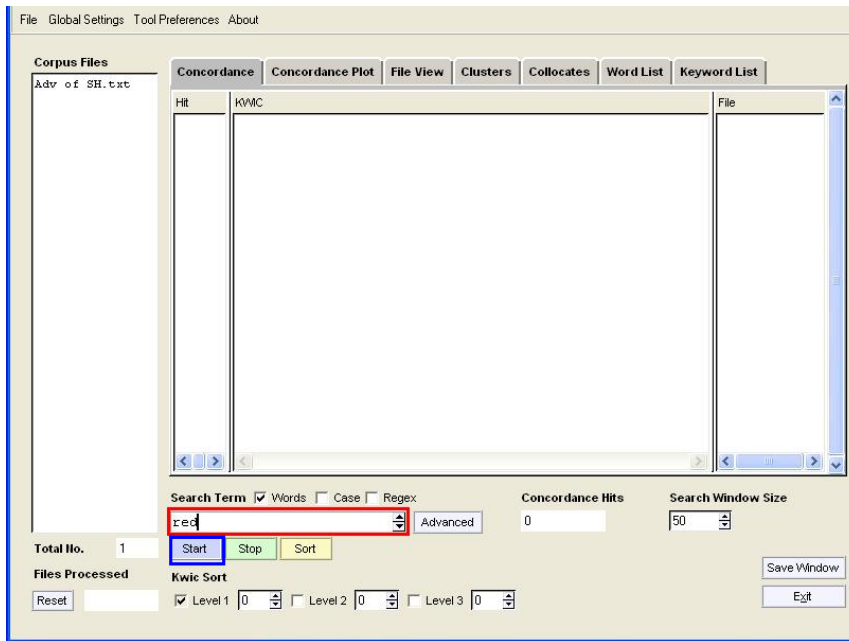
Corpus Files の下に分析したいテキストファイル名が出ていれば準備完了です。

## 2. 早速使ってみよう (KWIC 機能)

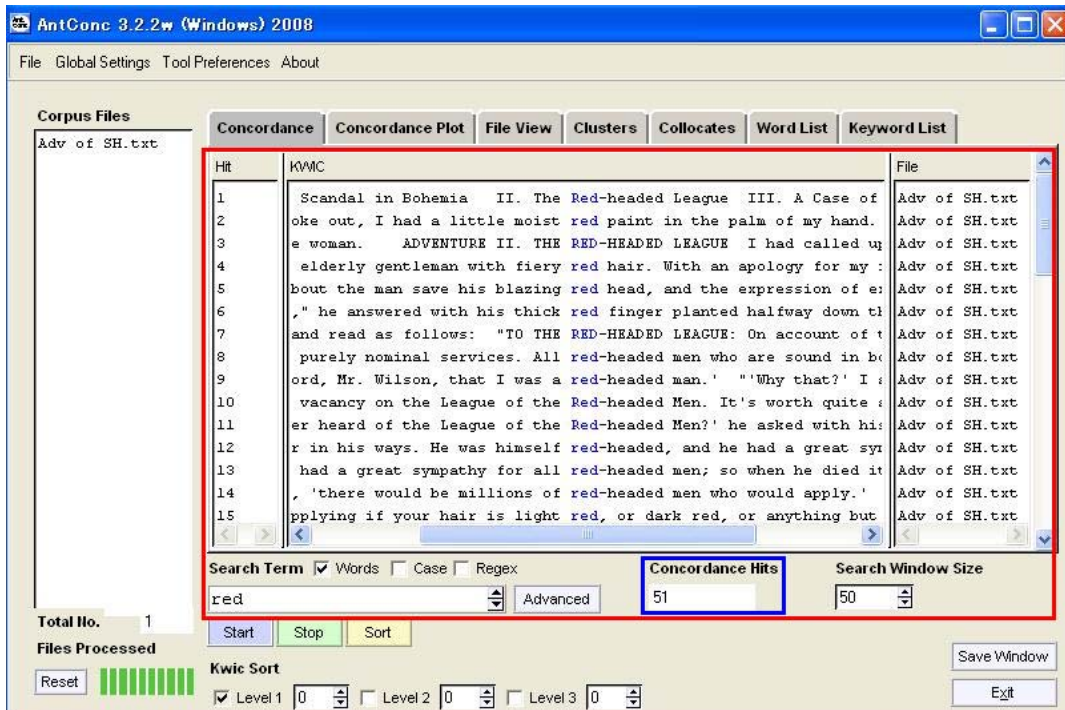
AntConc には様々な機能が実装されており、上の青枠内のタブを選んで機能を選択します。ソフトを実行すると最初は Concordance が選択されています。まずはこの“Concordance”機能を使ってみましょう。

Concordance 機能は別名 KWIC(Key Word In Context)とも呼ばれ、検索対象語（これを Node Word といいます）を含む行を取り出し、どのような前後関係の中で使われているかを一覧表示する機能です。

Search Term の中に調べたい単語を入れて、Start ボタンを押せば分析開始です。ここでは例として *The Adventure of Sherlock Holmes* (邦題『シャーロック・ホームズの冒険』) を用いて、”red”という単語で検索してみます。<sup>2</sup>



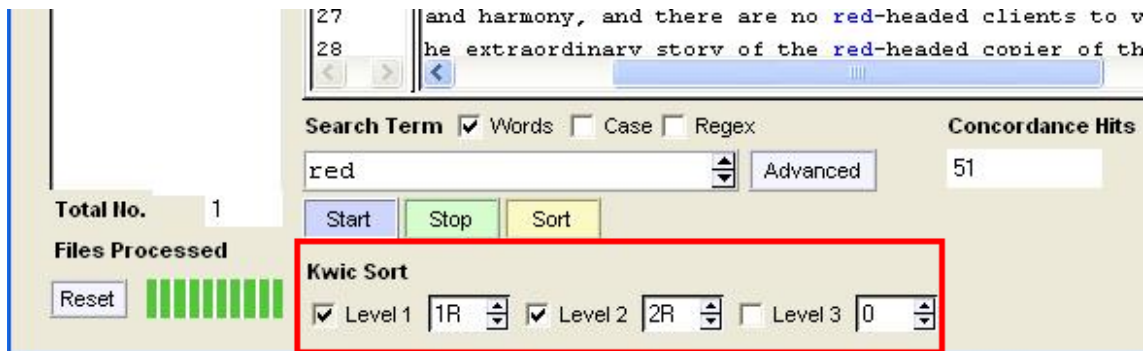
“Red”を含む文が 51 抽出できました！（red を中心に前後が見えます）



<sup>2</sup> “Project Gutenberg” [http://www.gutenberg.org/wiki/Main\\_Page](http://www.gutenberg.org/wiki/Main_Page) より無料で入手可能。

### 3. 出てきたものを並び替える SORT 機能

欲しい文を取り出すことはできましたが、このままだと調べたい特徴を見るのが大変かもしれません。そのために AntConc にはソート機能（並べ替え）があります。



上の KWIC Sort という部分で並び替えが可能です。

初期設定では [0] に設定してありますが、▲ボタンを押すと 1R, 2R...、▼ボタンを押すと 1L, 2L ... と数値が変化します。

これらは、検索語(Node Word)から見て、左もしくは右の何語目を基準に並び替えるか、ということを表します。R が右, L は左で、前の数字が何語目かということを表します。

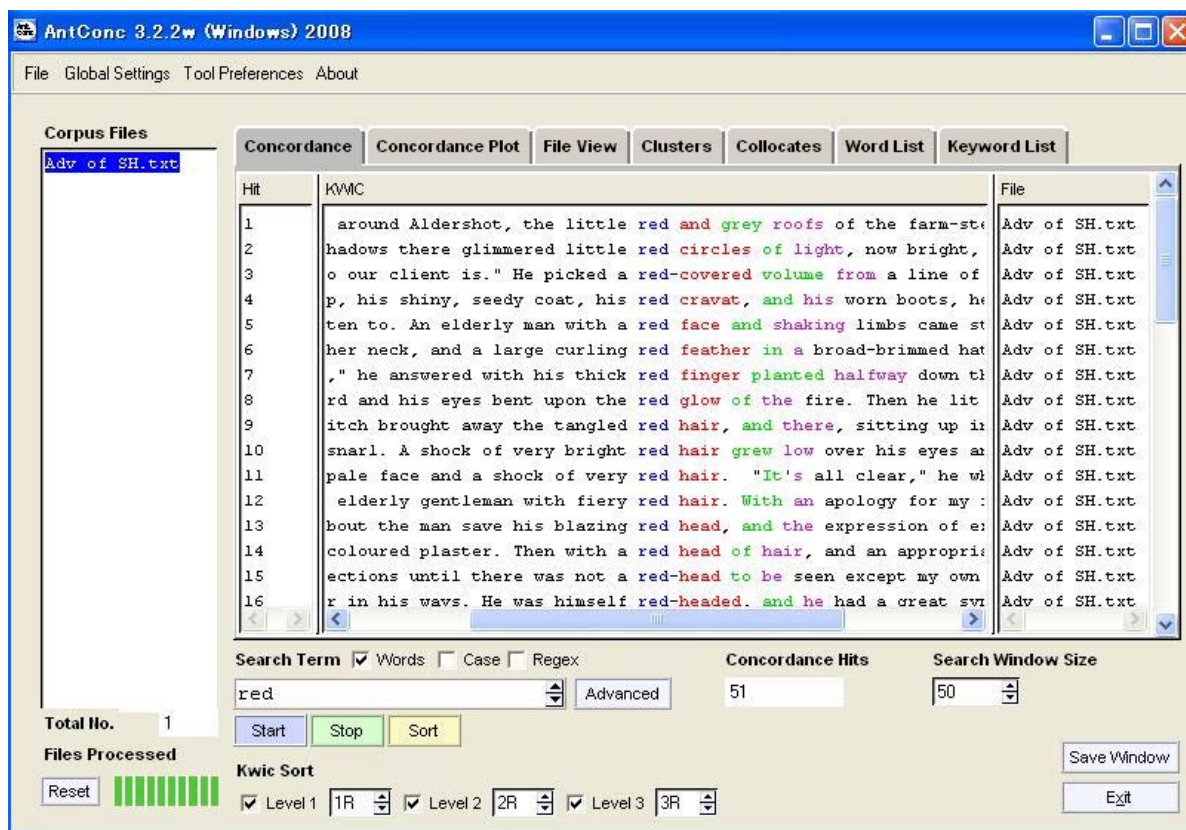
```
A shock of very bright red hair grew low over his  
5L 4L 3L 2L 1L NW 1R 2R 3R 4R 5R
```

左は red の前に来る語、右は red の後に来る語ですね。例えば 1R, 2R, 3R と入力した場合、まず red の直後に来る語(1R)で並び替えを行い、次に同じ”red hair”でも、その次に来る語(2R)を基準に、その後はそのまた次に来る語(3R)を基準に、と3段階で並び替えがなされます。(チェックボックス[ ]に、チェックを入れない限り機能しないので注意)

準備が終わったら、Sort ボタンを押します。



小さくて見づらいですが、並び替えられました。  
それぞれ Node Word が青、1R が赤、2R が黄緑、3R が紫で表示されています。



この機能を使用して並び替えることで、「後に何が来るか」「前に何が来るか」などの関係が見やすくなります。

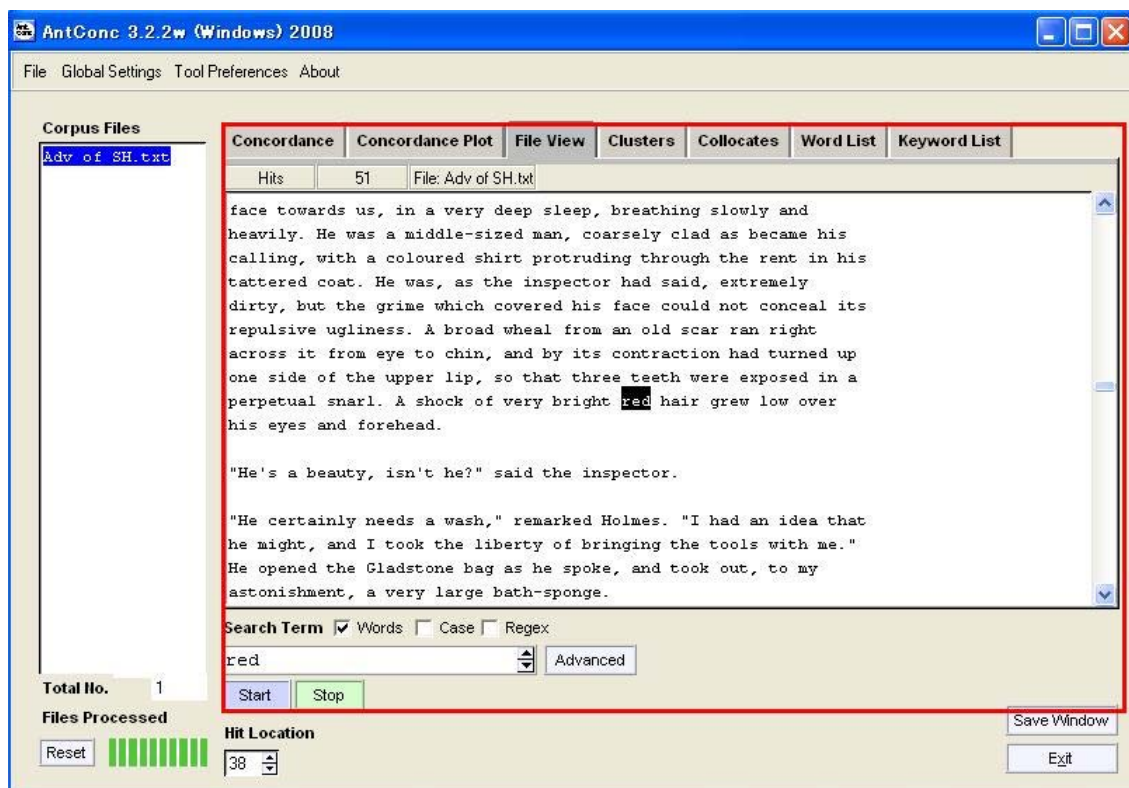


調べているうちに、表示される前後関係だけでは意味を把握できないときがあります。そんなときは、検索語(Node Word)をクリックしてみましょう。

( [ニア]のような手のマークになるはずです)

Hit	KVMC	File
1	around Aldershot, the little red and grey roofs of the farm-st	Adv of SH.txt
2	hadows there glimmered little red circles of light, now bright,	Adv of SH.txt
3	o our client is." He picked a red-covered volume from a line of	Adv of SH.txt
4	p, his shiny, seedy coat, his red cravat, and his worn boots, he	Adv of SH.txt
5	ten to. An elderly man with a red face and shaking limbs came st	Adv of SH.txt
6	her neck, and a large curling red feather in a broad-brimmed hat	Adv of SH.txt
7	, " he answered with his thick red finger planted halfway down th	Adv of SH.txt
8	rd and his eyes bent upon the red glow of the fire. Then he lit	Adv of SH.txt
9	itch brought away the tangled red hair, and there, sitting up in	Adv of SH.txt
10	snarl. A shock of very bright red hair grew low over his eyes an	Adv of SH.txt
11	pale face and a shock of very red hair. "It's all clear," he wh	Adv of SH.txt
12	elderly gentleman with fiery red hair. With an apology for my :	Adv of SH.txt
13	bout the man save his blazing red head, and the expression of e:	Adv of SH.txt
14	coloured plaster. Then with a red head of hair, and an appropri:	Adv of SH.txt
15	ections until there was not a red-head to be seen except my own	Adv of SH.txt
16	r in his ways. He was himself red-headed. and he had a great svr	Adv of SH.txt

すると、表示されている文章が、実際に分析対象のテキストで使われているところまで、本文にジャンプできます。(検索語は青色になっています)



#### 4. 検索に役立つ 正規表現 (Regular Expression) について

やや高度な検索方法として、正規表現による検索があります。正規表現とはいくつかの文字列を代数的に表現する方法で、例えば以下のようなものがあります。

**\* 直前の文字を 0 回以上繰り返す (アスタリスク)**

例) shoo\*t は shot にも shoot にも shooooot にもマッチします

**+ 直前の文字を 1 回以上繰り返す (プラス)**

例) shoo+t は shoot や shooooot にはマッチしますが shot にはマッチしません

**? 直前の文字が 0 回または 1 回ある (クエスチョンマーク)**

例) books? は book にも books にもマッチします

**. 任意の一文字 (ピリオド)**

例) s.t は sit にも sat にも set にもマッチします

**¥w 英数字あるいはアンダーバーが一文字 (円記号とダブリュー)**

例) ¥w+ は任意の英数文字列にマッチします

**¥d 数字一文字 (円記号とディー)**

例) ¥d+ は数字列にマッチします

**| 選択 (縦棒)**

例) is|am|are は is、am、are のいずれかにマッチします

**() グループ化 (かっこ)**

例) s(al)it は sat あるいは sit にマッチします

**[] 中の一文字にマッチする (角かっこ)**

例) s[ai]t は sat あるいは sit にマッチします

**^ 行頭 (ハット)**

例) ^I は行頭の I のみにマッチします

**\$ 行末 (ドル記号)**

例) me\$ は行末の me のみにマッチします

**[^] ~以外 (角かっことハット)**

例) s[^e]t は sat や sit にはマッチしますが、set にはマッチしません

**¥s 空白文字、つまりスペースやタブ (円記号とエス)**

例) ¥w+¥s¥w+ は任意の二単語にマッチします

**¥b 単語区切り (円記号とビー)**

例) in は into や sin にもマッチしますが、¥bin¥b は in にしかマッチしません

**¥ エスケープ文字 (上記の記号を文字として処理する) (円記号)**

例) ピリオドを検索したい場合は ¥. とします

## 5. タグセット一覧とその活用法

以下が今回使用するタグの一覧です。

---

CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NP	Proper noun, singular
NPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PP	Personal pronoun
PPS	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	to
UH	Interjection
VB	Be-verb, base form
VBD	Be-verb, past tense
VBG	Be-verb, gerund or present participle
VBN	Be-verb, past participle
VBP	Be-verb, non-3rd person singular present
VBZ	Be-verb, 3rd person singular present
VH	Have, base form
VHD	Have, past tense
VHG	Have, gerund or present participle
VHN	Have, past participle
VHP	Have, non-3rd person singular present
VHZ	Have, 3rd person singular present
VV	Other verbs, base form
VVD	Other verbs, past tense
VVG	Other verbs, gerund or present participle

VVN	Other verbs, past participle
VVP	Other verbs, non-3rd person singular present
VVZ	Other verbs, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP\$	Possessive wh-pronoun
WRB	Wh-adverb

---

タグは「表層形 (タブ) タグ (タブ) レンマ」という形式で、一語一行で付与されています。例えば「There are many problems facing the Monarchy.」という文であれば、以下のようになります。

There	EX	There
are	VBP	be
many	JJ	many
problems	NNS	problem
facing	VVG	face
the	DT	the
Monarchy	NN	monarchy
.	SENT	.

AntConc では行末の改行には¥s がマッチします。そのことと正規表現を組み合わせることにより、様々な項目が抽出可能となります。

### 進行相

→be 動詞+ -ing 形の動詞

→レンマが be の語+タグが V[何か一文字]G

→例えば以下のような形

is VBZ be

eating VVG eat

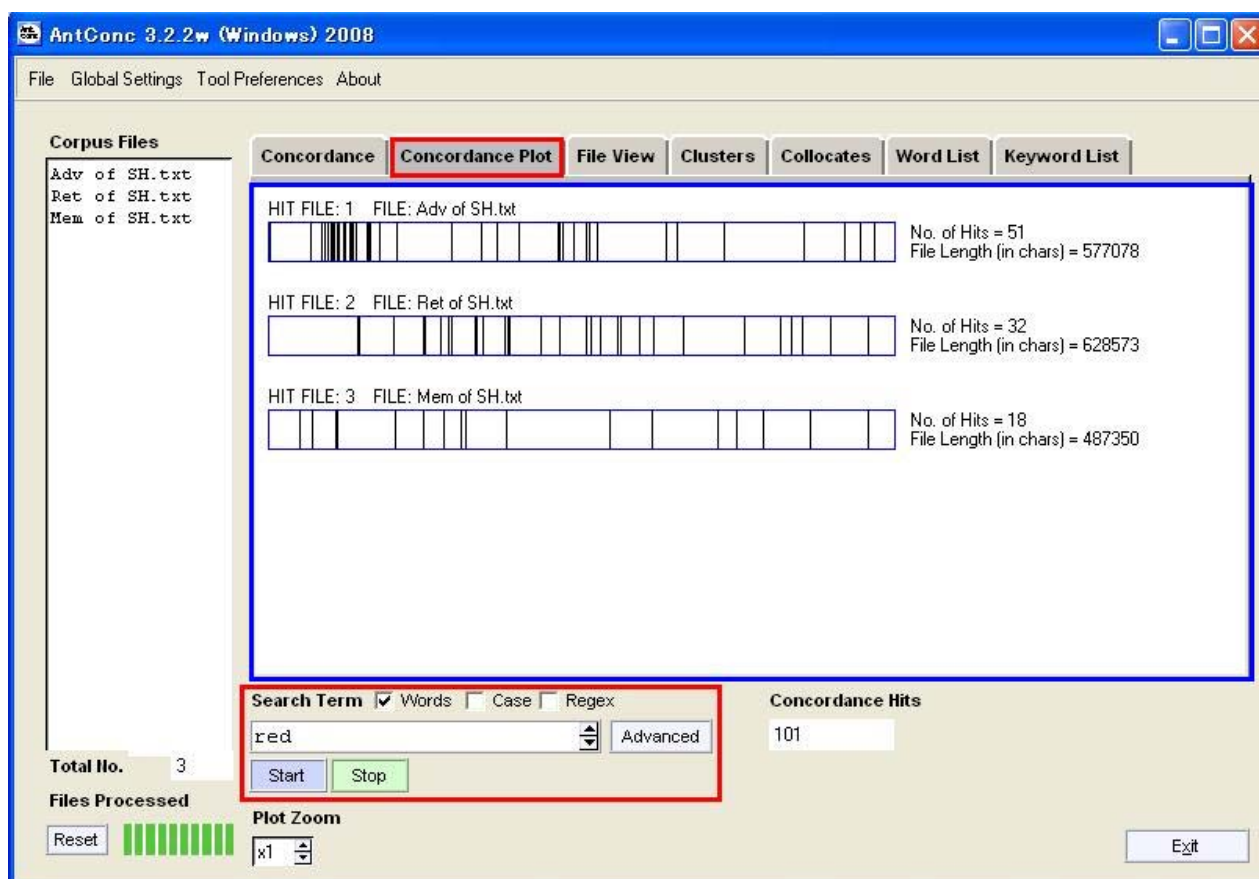
→be 動詞 : ¥bbe¥s -ing 形の動詞 : ¥w+¥sV.G

→¥bbe¥s¥w+¥sV.G

## 6. 単語の分布を調べる Concordance Plot 機能

検索した語がテキスト中、テキスト間でどのように分布しているかを調べたい場合には、左から2番目の Concordance Plot タブをクリックして選択します。

Concordance で調べた人はそのまま、調べていない人は下のボックスに検索したい語を入れます。ここでは *The Adventure of Sherlock Holmes* に加えて、*The Memoir of Sherlock Holmes*(邦題『シャーロック・ホームズの回想』)、*The Return of Sherlock Holmes*(邦題『シャーロック・ホームズの帰還』)の3テキストを用いました。



すると、上の青枠のような結果が得られます。

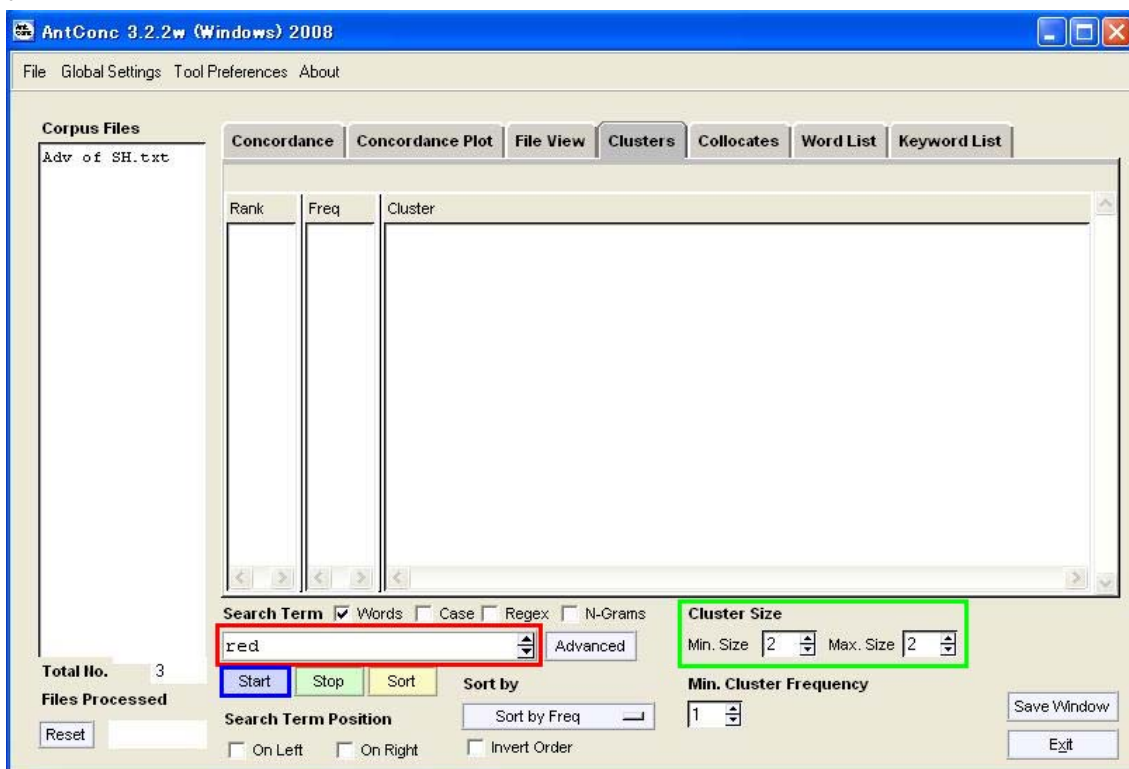
横長のバーが各テキストファイル、縦線の入っている位置に検索語 red が出現しているという意味です。上を見ると、red は『冒険』の、しかも一部に集中して使用されているらしいことが分かります。(“The Red-headed League”、邦題『赤毛組合』)のトピックのためだと推測されます)

## 7. まとまった語句を調べる Clusters 機能

2章で、調べたい単語を含む文脈を検索しました。

一つ一つ文を見ることそれ自体は貴重な情報をもたらしてくれますが、一方で非常に時間と労力を必要とする欠点を持ちあわせています。

そこで、検索語を含む「かたまり」(chunk, cluster などと呼ばれます) で数値として取り出す Clusters 機能を使うことで、頻度を数値で取り出すことができます。



まずは **Concordance** のときと同じく、検索したい語を赤で囲ってあるボックスに入れます。

次に、緑で囲った **Cluster Size** を指定します。これはそのまま、**Min. Size** が「最小単位」、**Max. Size** が「最大単位」、つまり何語から何語までの大きさをかたまりとみなすか、という設定です。例えば、**Min** を 2, **Max** を 4 と入力すると、2語から 4 語、つまり 2 語かたまり、3 語かたまり、4 語かたまりの語を **cluster** とみなして分析、出力します。

Concordance | Concordance Plot | File View | Clusters | Collocates | Word List | Keyword List

Total No. of Cluster Types: 164 | Total No. of Cluster Tokens: 227

Rank	Freq	Cluster
1	11	red-headed
2	6	Red-headed
3	5	a red
4	5	the red
5	4	red hair
6	4	Red-headed League
7	4	the Red
8	4	the Red-headed
9	3	of red
10	3	of the Red
11	3	RED-HEADED
12	3	RED-HEADED LEAGUE
13	3	red-headed men
14	3	THE RED

Search Term  Words  Case  Regex  N-Grams

red

Cluster Size Min. Size  Max. Size

Sort by

Search Term Position

Min. Cluster Frequency

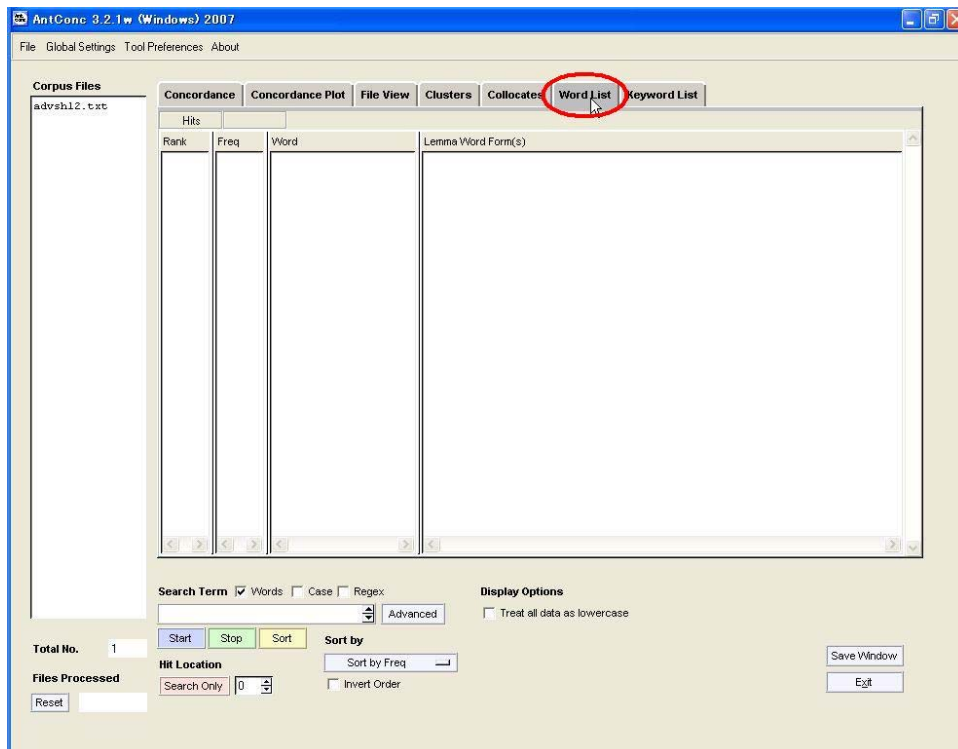
On Left  On Right  Invert Order

すると上のように、検索語を含むかたまりを自動で計算してくれ、頻度に基づいた情報が得られます。

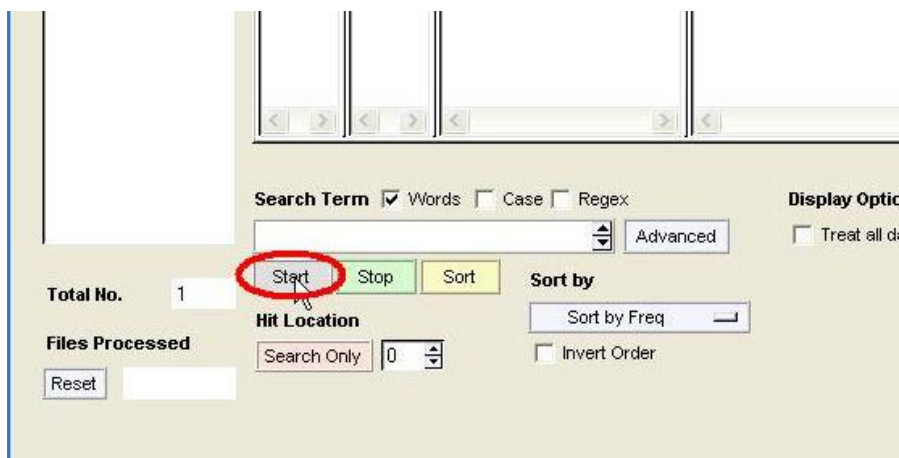
## 8. Word List 機能の使い方

Word List 機能を使えば、対象ファイル中の語を頻度順・アルファベット順に見ることができます。手順は以下の通りです。

Word List のタブをクリックします



まずはそのまま左下の **Start** をクリックします





すると処理対象のファイルの中に出現する語が頻度順に表示されます

The screenshot shows a software interface for processing text files. On the left, a list of 'Corpus Files' includes 'advsh12.txt'. The main window displays a table of words sorted by frequency. The table has columns for Rank, Frequency, and Word. The words listed are: the (5340), I (3038), and (2863), to (2775), of (2678), a (2568), in (1717), that (1661), was (1392), it (1308), you (1278), he (1165), is (1127), his (1103), have (911), my (907), had (821), with (811), as (780), which (765), at (738), for (712), and be (643). The table is highlighted with a red border. Below the table, there are search options: 'Search Term' with checkboxes for 'Words', 'Case', and 'Regex'. There are also buttons for 'Start', 'Stop', and 'Sort'. The 'Sort by' dropdown is set to 'Sort by Freq'. The 'Hit Location' section shows 'Search Only' and a value of '0'. The 'Total No.' is '1' and 'Files Processed' is '0'.

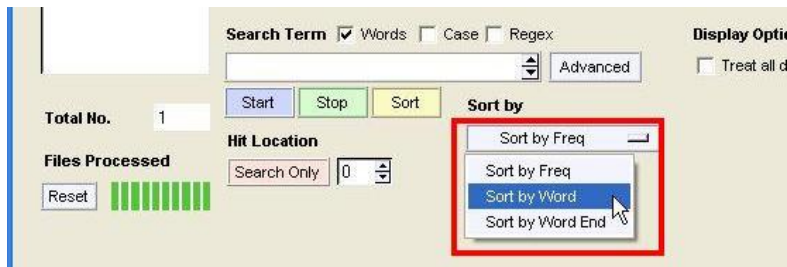
Rank	Freq	Word
1	5340	the
2	3038	I
3	2863	and
4	2775	to
5	2678	of
6	2568	a
7	1717	in
8	1661	that
9	1392	was
10	1308	it
11	1278	you
12	1165	he
13	1127	is
14	1103	his
15	911	have
16	907	my
17	821	had
18	811	with
19	780	as
20	765	which
21	738	at
22	712	for
23	643	be

語をクリックするとその語を KWIC 形式で見ることができます

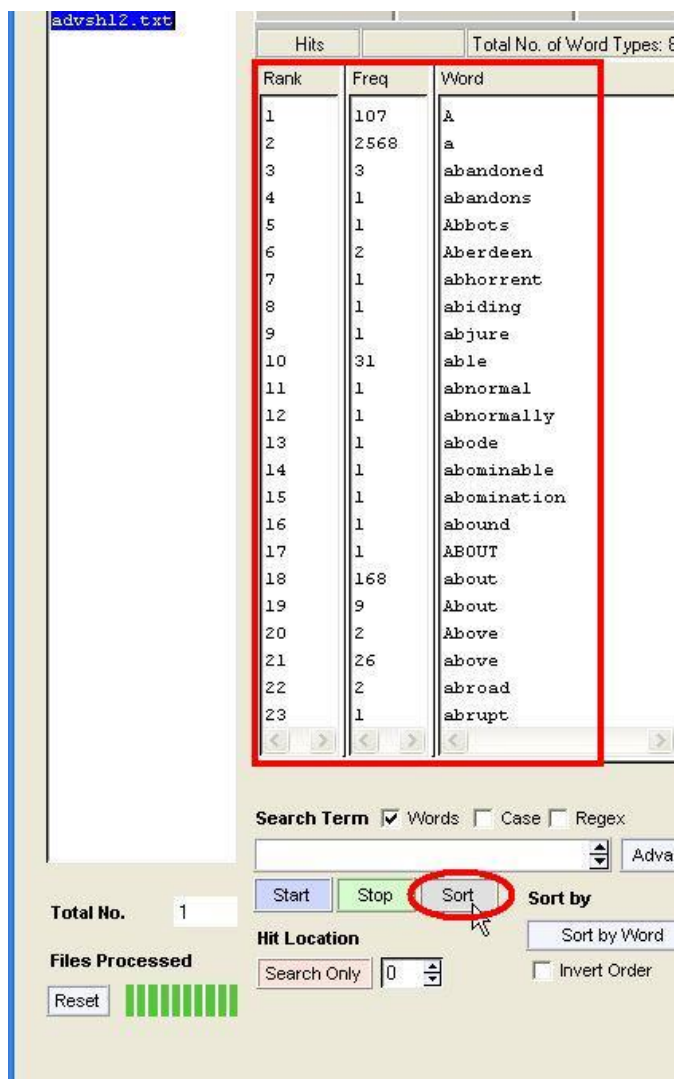
This screenshot is identical to the one above, but with the word 'have' in the 'Word' column of the table highlighted with a red box. This indicates that the user has clicked on the word, and the software is now displaying the KWIC (Key Word In Context) view for that word. The interface elements below the table remain the same.

Rank	Freq	Word
1	5340	the
2	3038	I
3	2863	and
4	2775	to
5	2678	of
6	2568	a
7	1717	in
8	1661	that
9	1392	was
10	1308	it
11	1278	you
12	1165	he
13	1127	is
14	1103	his
15	911	have
16	907	my
17	821	had
18	811	with
19	780	as
20	765	which
21	738	at
22	712	for
23	643	be

アルファベット順でソートしてみます。まずは Sort by Word をクリックします



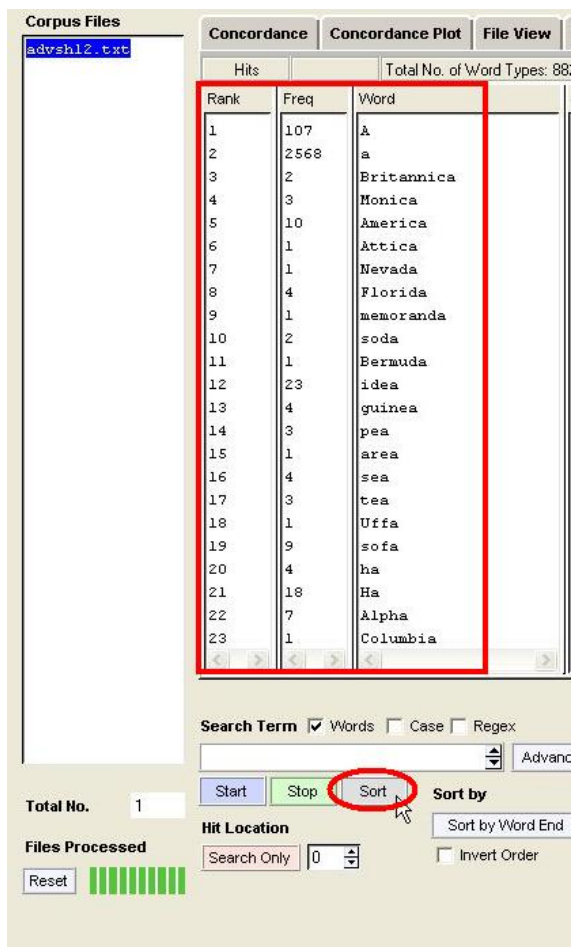
Sort をクリックすると、アルファベット順でソートされました



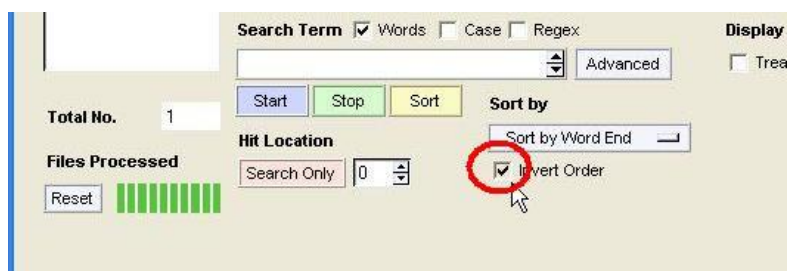
今度は語末のアルファベット順でソートしてみます。Sort by Word End を選択します。



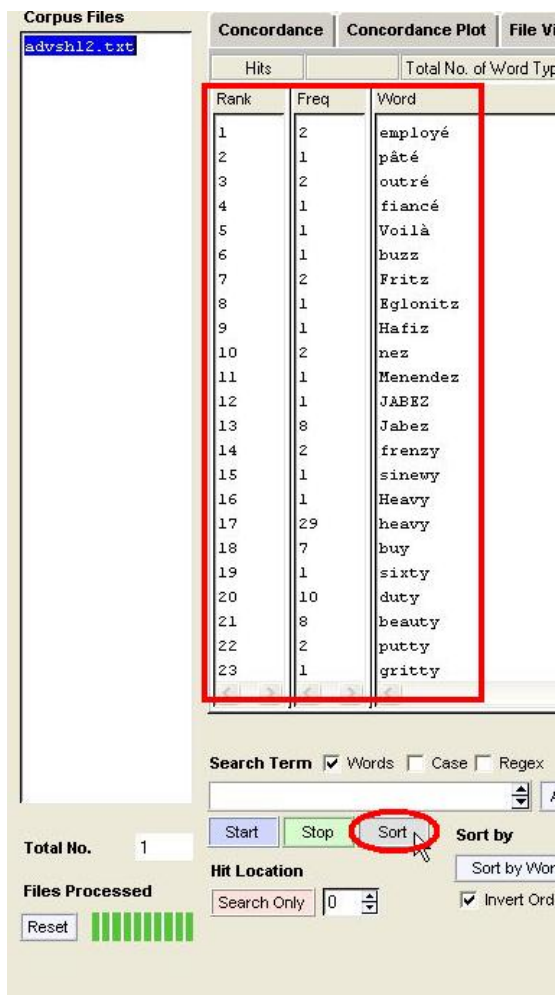
ソートをクリックすると、語末のアルファベット順でソートされました。



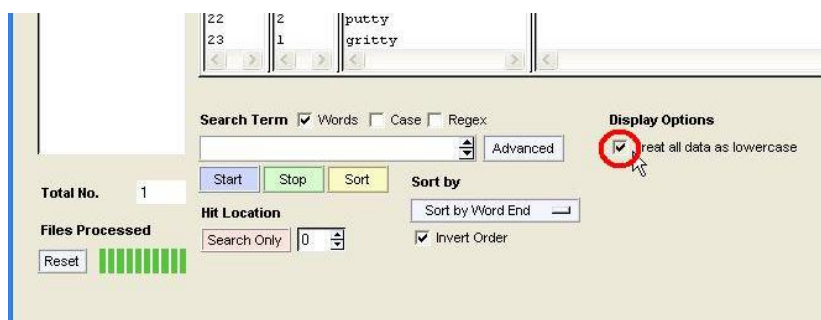
語末を逆アルファベット順にしてみます。Invert Order にチェックを入れます。



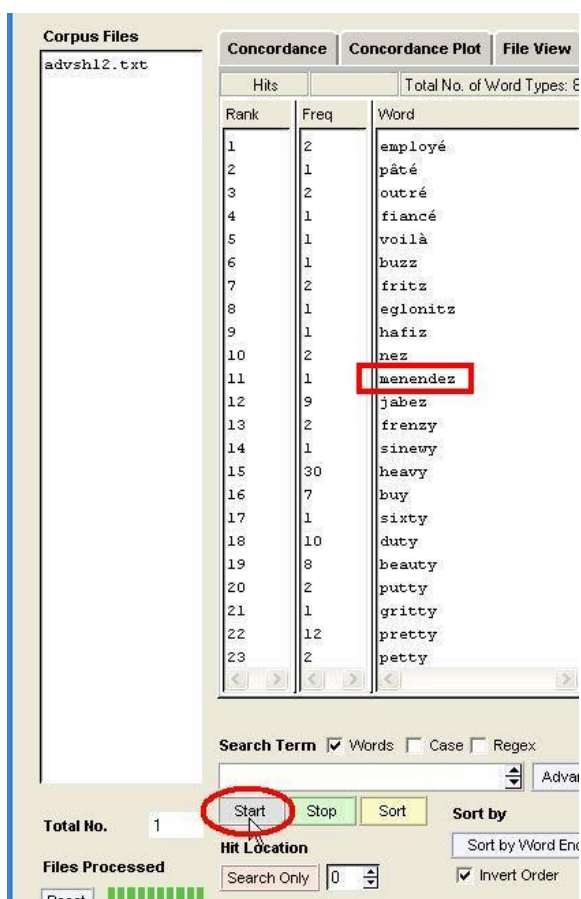
Sort をクリックすると、語末が逆アルファベット順になりました



大文字/小文字の区別をなくしたい場合は **treat all data as lowercase** にチェックを入れます



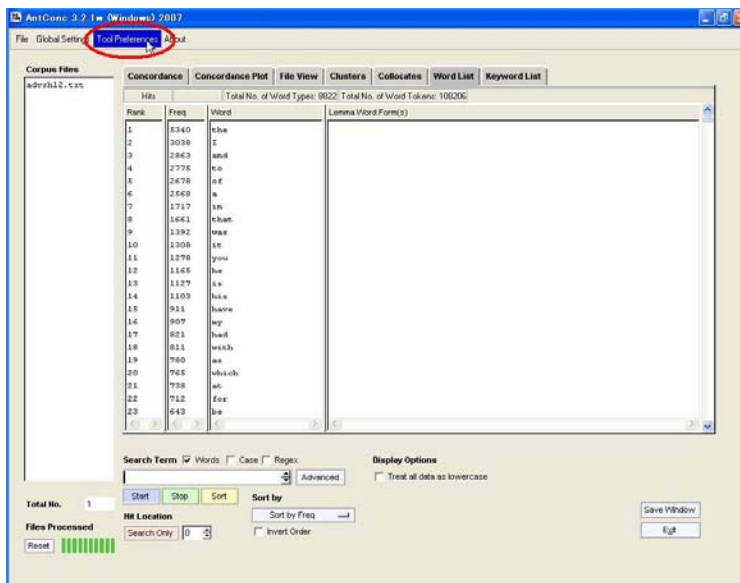
今度は **Sort** ではなく **Start** をクリックします。元々大文字だった文字も小文字として扱われるようになりました。



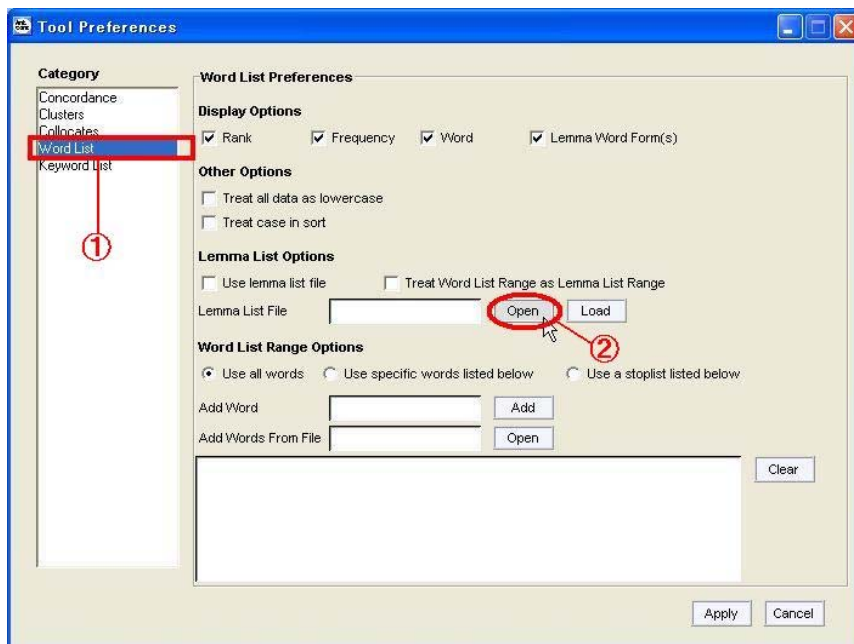
## 9. レンマ化(Lemmatization)の方法

上記手法では、語形が異なっていれば全て別の語としてカウントされてしまいます。つまり **have**、**has**、**had**、**having**、**'ve** は全て別の語として扱われます。これらを全て同一の語として扱うには、リストをレンマ化（レンマ：辞書の見出しの形）する必要があります。以下ではその方法を説明します。

画面上部の Tool Preferences をクリックします

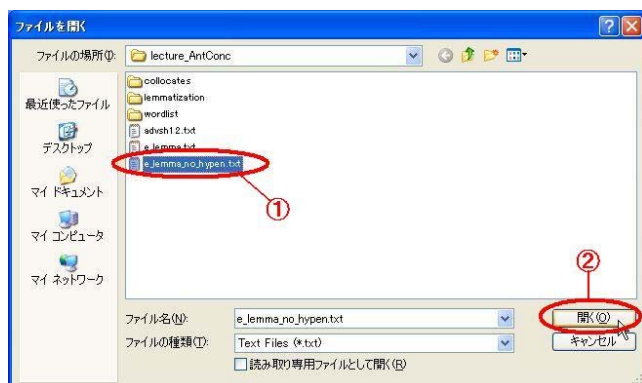


Category 下の Word List を選択し、Lemma List Options という項目内にある Open をクリックします

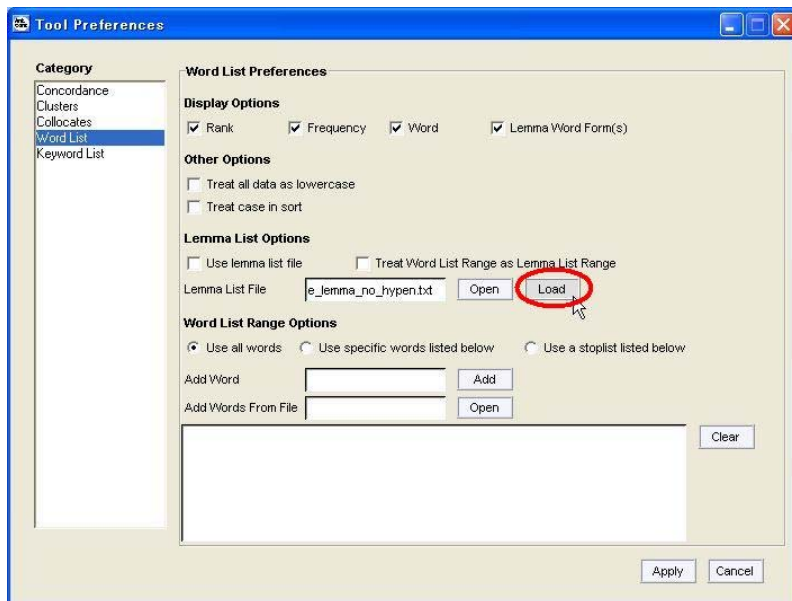


e\_lemma\_no\_hyphen.txt を選択します

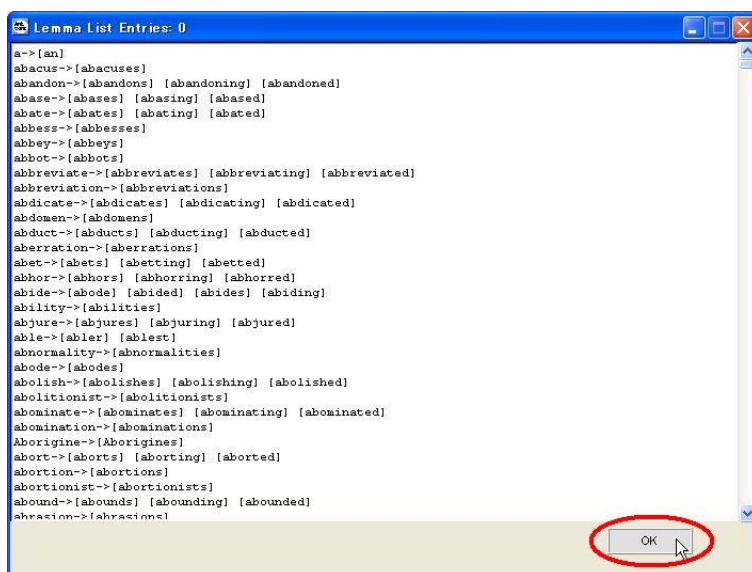
(e\_lemma\_no\_hyphen.txt は [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html) から入手できます。 e\_lemma.txt というファイルもありますが、うまく動作しないためハイフンを含まないこちらのファイルの方が良いと思われます。)



Load をクリックします

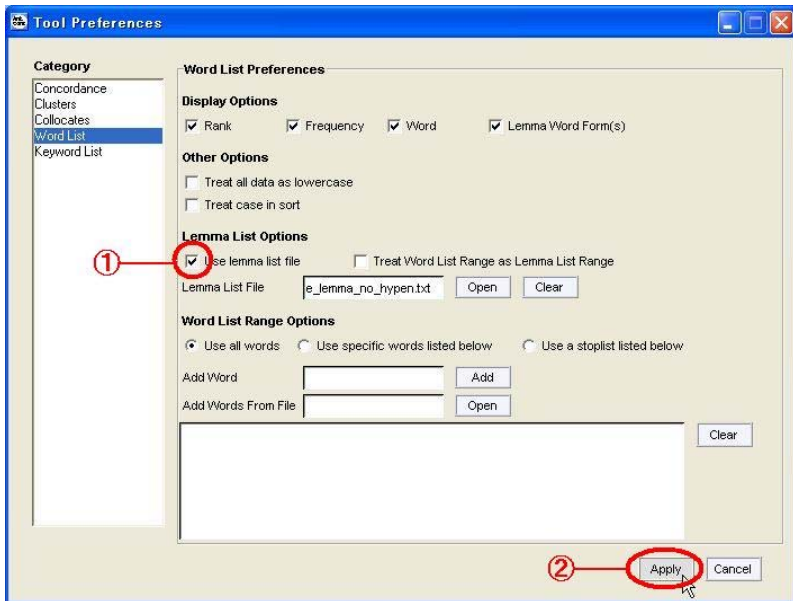


下のようなウィンドウが現れますので、OK をクリックします





Use lemma list file にチェックを入れ、Apply をクリックします



Start をクリックすると、レンマ化された Word List が出てきます。be というレンマの中に am、are、be、been などの語が全て含まれているのがわかります。

Rank	Freq	Lemma	Lemma Word Form(s)
1	5340	the	
2	4438	be	am 185 are 346 be 643 been 395 is 1127 m 4 was 1392 were 346
3	3038	I	
4	2894	a	a 2568 an 326
5	2863	and	
6	2775	to	
7	2678	of	
8	2086	have	had 821 has 282 have 911 having 58 ve 14
9	1724	that	that 1661 those 63
10	1717	in	
11	1308	it	
12	1278	you	
13	1165	he	
14	1103	his	
15	907	my	
16	811	with	
17	780	as	
18	765	which	
19	738	at	
20	712	for	
21	635	me	
22	625	not	
23	595	say	said 486 say 94 says 15

Search Term:  Words  Case  Regex  Advanced  Treat all data as lowercase

Display Options:  Treat all data as lowercase

Start Stop Sort Sort by

Sort by Freq

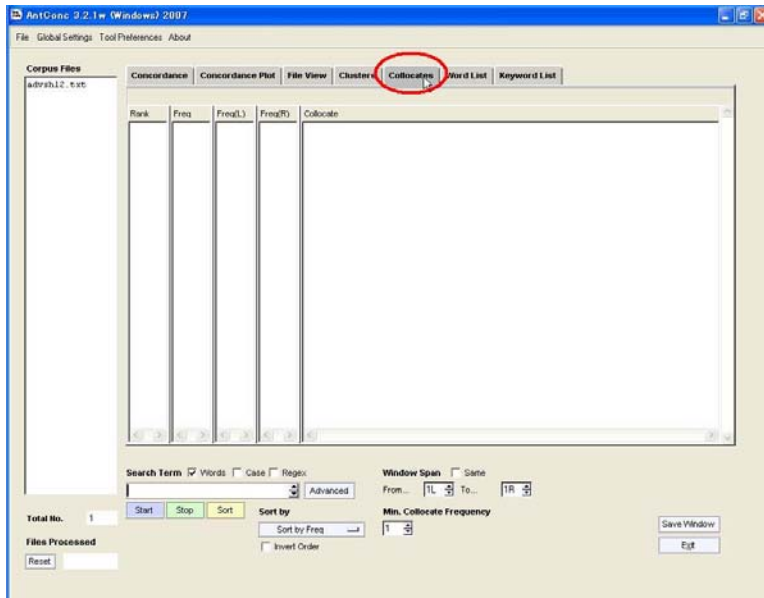
Search Only: 0

Invert Order

## 1 0 . Collocates 機能の使い方

Collocates 機能では、ファイル中の語の共起関係を見ることができます。手順は以下の通りです。

Collocates タブをクリックします



Search Term に Holmes と入力し、Start をクリックします



Holmes という語と共起する語が出てきました。頻度順になっており、左から順に順位、総頻度、対象語 (Holmes) の左側に出現する頻度、対象語の右側に出現する頻度となっています。Collocate 欄の語をクリックすれば、その語を KWIC 形式で見ることができます。

Rank	Freq	Freq(L)	Freq(R)	Collocate
1	460	0	0	Holmes
2	119	111	8	said
3	96	96	0	Sherlock
4	70	70	0	Mr
5	31	0	31	I
6	24	3	21	and
7	16	0	16	had
8	15	2	13	that
9	15	3	12	as
10	13	0	13	was
11	13	12	1	asked
12	12	11	1	remarked
13	10	4	6	when
14	9	0	9	the
15	8	2	6	with
16	8	0	8	sat
17	8	0	8	laughing
18	7	3	4	you
19	7	0	7	It
20	7	2	5	in
21	6	4	2	to
22	6	0	0	HOLMES
23	6	3	3	but

Search Term:  Words  Case  Regex  
 Holmes  
 Window Span: From... 1L  
 Sort by: Sort by Freq  
 Min. Collocate: 1  
 Invert Order

先ほどは Holmes の直前と直後に来る語しか見ていませんでしたが、今度は共起関係を判断する枠を広げてみます。Window Span をそれぞれ 5L、5R に設定します。これは共起関係を判断する対象として、対象語 (Holmes) の 5 個左の語 (5L) から 5 個右の語 (5R) までを見るということです。

Window Span:  Same  
 From... 5L To... 5R  
 Min. Collocate Frequency: 1  
 Invert Order

Start をクリックすると、先ほどとは少し異なる結果が出てきました

The screenshot shows the Concordance tool interface. At the top, there are tabs for 'Concordance', 'Concordance Plot', 'File View', and 'Clusters'. Below the tabs, it displays 'Total No. of Collocate Types: 1348' and 'Total No. of Collocate Tokens: 5121'. A table lists collocates with columns for Rank, Freq, Freq(L), Freq(R), and Collocate. The 'Start' button is circled in red.

Rank	Freq	Freq(L)	Freq(R)	Collocate
1	462	1	1	Holmes
2	188	80	108	the
3	136	123	13	said
4	134	40	94	I
5	98	45	53	to
6	97	96	1	Sherlock
7	91	88	3	Mr
8	88	21	67	and
9	81	49	32	you
10	76	58	18	of
11	73	34	39	a
12	67	11	56	his
13	66	26	40	that
14	59	25	34	in
15	55	19	36	is
16	42	13	29	had
17	41	16	25	have
18	39	21	18	was
19	39	23	16	my
20	39	24	15	it
21	34	10	24	as
22	33	14	19	with
23	30	13	17	at

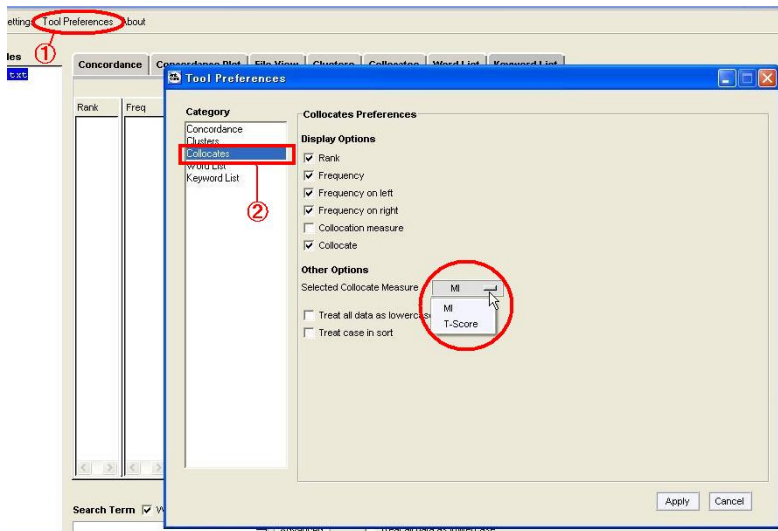
Search Term:  Words  Case  Regex  
 Holmes  
 Start Stop Sort  
 Sort by: Sort by Freq  
 Invert Order

Sort by は Word List 機能とほぼ同じです。Min. Collocate Frequency では頻度が一定値以下の語を対象外にすることができます。

The screenshot shows the 'Sort by' and 'Min. Collocate Frequency' options highlighted with a red box. The 'Sort by' dropdown is set to 'Sort by Freq' and the 'Min. Collocate Frequency' is set to 1.

Search Term:  Words  Case  Regex  
 Holmes  
 Start Stop Sort  
 Sort by: Sort by Freq  
 Min. Collocate Frequency: 1  
 Invert Order

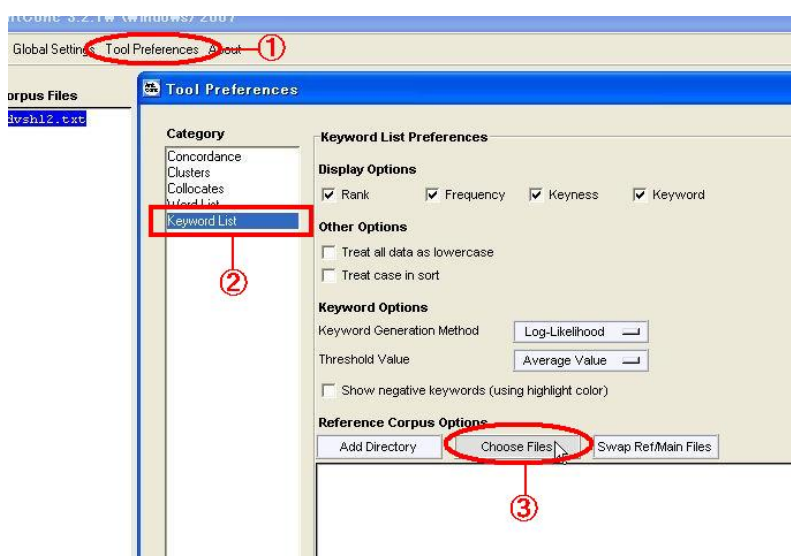
他のオプションは Tool Preferences → Collocates で見るすることができます。共起関係を算出する統計手法はデフォルトでは MI スコアを用いていますが、T スコアを用いることもできます。



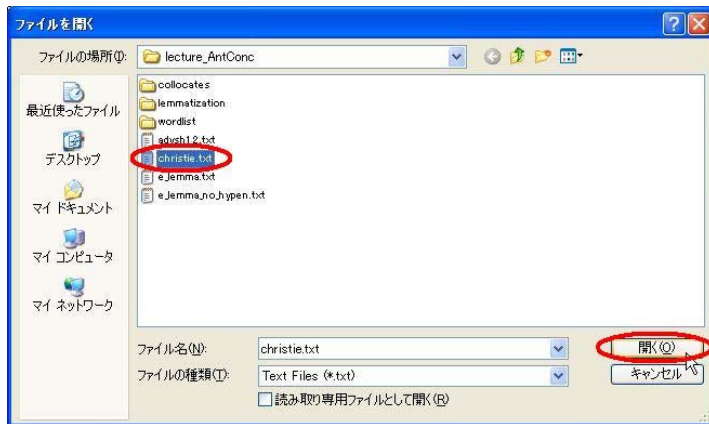
## 1 1 . Keyword List 機能

Keyword List 機能では、二つのテキストを比較し、一方のテキストに頻度面で特徴的な語を抽出することができます。ここではこれまで用いてきたコナンドイルの文章をアガサクリスティーの文章と比較することにより、双方の特徴的な語彙を明らかにします。

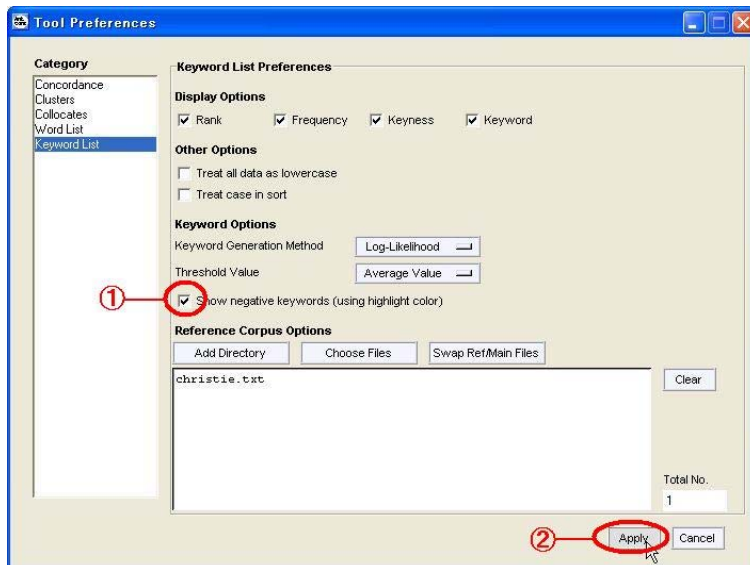
Tool Preferences→Keyword List と進み、Reference Corpus Options というところの Choose Files をクリックします



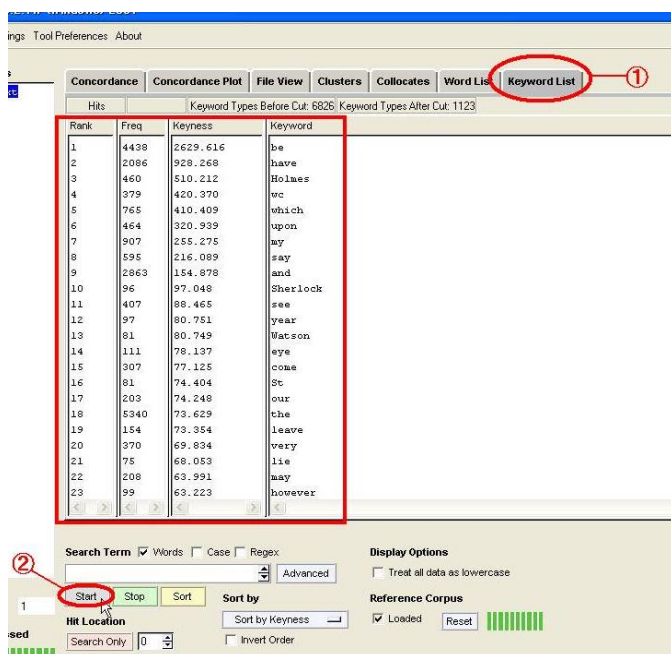
アガサクリスティーのファイルを選択します



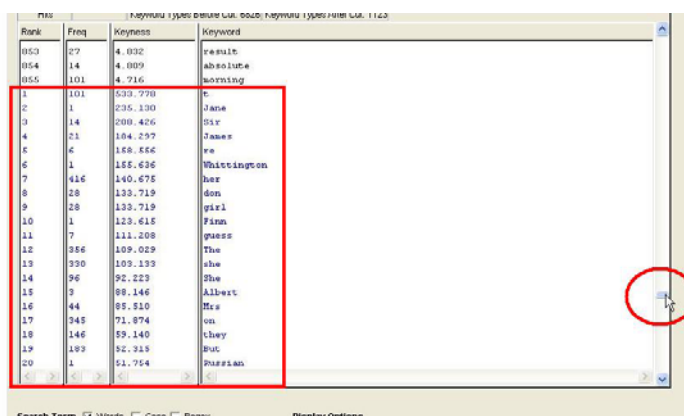
コナンドイルの文章だけでなくアガサクリスティーの文章の特徴的な語も見たい場合は、**Show negative keywords** のチェックを入れます。その後、**Apply** をクリックします。



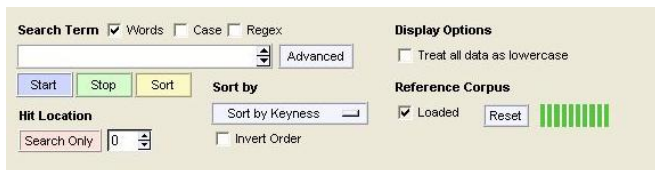
Keyword List タブをクリックし、Start をクリックすると、まずはコナンドイルの文章に特徴的な語を見ることができます。Keyness とはもう一方のテキストと比較し、どの程度高頻度であるかを示す指標です。



スクロールバーを下げていくと、青い文字が見えます。これがアガサクリステイの文章に特徴的な語です。



画面下部のオプションは Word List 機能と共通です



Tool Preferences→Keyword List で他の設定を行うことができます。Keyness の算出手法等もここで変更できます。

