

AntConc (Windows, MacOS, Linux)

Build 4.0.8

Laurence Anthony, Ph.D.

Center for English Language Education in Science and Engineering, School of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan April 20, 2022

arget Corpus lame: AmE06_Learned iles: 80		WIC Plot al Hits: 42 Page	File Cluster N-Gram Collocate Word Keyword e Size 100 hits V I to 42 of 42 hits Image: Collocate Image: Collocate <th></th> <th></th>						
okens: 161469		File	Left Context	Hit	Right Context				
AmE06_J01.txt	1	AmE06_J32.txt	ing word is a verb, but not when the -ing	word	is a noun or adjective. So the grammar of				
AmE06_J02.txt AmE06_J03.txt	2	AmE06_J47.txt	bject, action or quality. Learning the category boundaries for each	word	is a specific "problem of induction." Children are placed				
AmE06_J04.txt	3	AmE06_J32.txt	a-hunting dog. The a- is possible when the - ing	word	is a verb, but not when the -ing word				
AmE06_J05.txt AmE06_J06.txt	4	AmE06_J34.txt	which coincides with word-initial position, and that the entire	word	is dominated by a single syllable. The gesture-calculations				
AmE06_J07.txt AmE06_J08.txt	5	AmE06_J59.txt	as Marjorie Perloff puts it, or perhaps that they put	word	and image in a mutually interpretive context, as the				
mE06_J09.txt	6	AmE06_J59.txt	Stein's verbal portraits of these painters "attempt to fuse	word	and image," as Marjorie Perloff puts it, or perhaps				
eference Corpus	7	AmE06_J60.txt	an ancient purity and directness, a pre-Babelic unity of	word	and thing. In so doing, though, photography was just				
ame: AmE06	8	AmE06_J47.txt	the word, or if the child understands and says the	word.	The second component is a list of 63 communicative, social,				
les: 500 ikens: 1017879	9	AmE06_J47.txt	to indicate if they have heard their child say the	word.	The second major component utilizes an innovative sentence pair				
mE06_A01.txt ^	10	AmE06_J47.txt	Consider what would appear to be the simplest condition of	word	learning. A fluent speaker of the language, such as				
mE06_A02.txt mE06_A03.txt	11	AmE06_J47.txt	MCDI and laboratory measures of comprehension, production, and	word	learning. The second is the Twins' Early Development Study (
mE06_A04.txt mE06_A05.txt	12	AmE06_J47.txt	Each type of reference is a plausible use of the	word.	And even if we were correct in believing that "				
mE06_A06.txt mE06_A07.txt	Sea	rch Query 🗹 V	Vords Case Regex Results Set All hits ~ Con	ntext Size	10 token(s)				
mE06_A08.txt	wo	ord	✓ Start	Adv Searc	h				

Introduction

AntConc is a freeware, multiplatform tool for carrying out corpus linguistics research, introducing corpus methods, and doing data-driven language learning. It runs on any computer running Microsoft Windows (built on Win 10), MacOS (built on Mac Catalina), and Linux (built on Linux Mint). It is developed in *Python* and *Qt* using the *PyInstaller* compiler to generate executables for the different operating systems. It uses SQLite as the underlying database.

Getting Started

💐 Windows - Installer

Double click the *AntConc.exe* file and follow the instructions to install the application into your Programs folder. You can delete the .exe file when you are finished. You can start the application via the Start Menu.

💐 Windows - Portable

Unzip the *AntConc.zip* file into a folder of your choice. In the *AntConc* folder, double click the *AntConc.exe* file to launch the program.

Macintosh OS X

Double click the *AntConc.dmg* file to create a AntConc disk image on your desktop. Open the disk image and drag and drop the AntConc app onto the Applications folder (or into another location if you desire). You can then launch the app by double clicking on the icon in the Applications folder or the Launchpad.

Å Linux

Decompress the *AntConc.tar.gz* file into a folder of your choice. In the *AntConc* folder, double click the *AntConc.sh* file to launch the software. On the command line, type ./AntConc.sh to launch the software.

Overview of Tools

KWIC Plot File Cluster N-Gram Total Hits: 77 Total Files With Hits: 20

AmE06 M7 tet 1993

AmE06_J32.txt 2073

AmE06 J75.txt 2041

AmE06_J05.txt 2025

AmE06_J44.txt 2027

mE06_J60.txt 2023

AmE06_J19.txt 2018

AmE06_J02.txt 2036

AmE06 J17.txt 2017

AmE06 J18.txt 1884

AmE06 J35.txt 2047

Sort by Dispersion V Invert Order

Search Query Vords Case Regex Results Set All hits

11

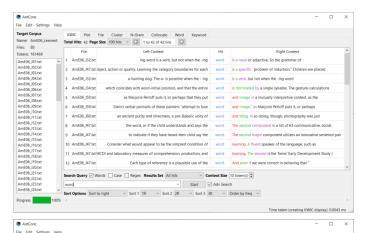
DocTokens Freq Nor

17 8529.854 0.638

DocID DocPath

larget Corpus lame: AmE06_

AntConc contains eight tools that can be accessed either by clicking on their 'tabs' in the tool window, using CTRL+TAB to toggle through the tools, or using the key combination CTRL + Tool Number (e.g., CTRL +1 for KWIC, CTRL +2 for Plot) to select a specific tool.



Collocate

5306.319 0.63 11

> 4409.603 0.613 0.521

3456.79

1973.36 0.447

1977.261

2477.701

982.318 0.333

991 572 0 333

1061.571 0.333

0.333

Start Adv

V Plot Zoom 1.00 x Color

2 977.04

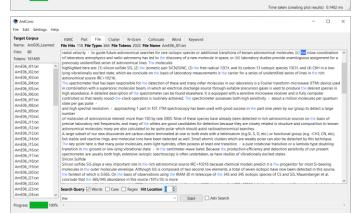
KWIC (Key-Word-In-Context) Tool

This tool shows search results in a concordance or 'KWIC' (Key-Word-In-Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

Plot Tool

File Tool

This tool shows concordance search results plotted in a 'barcode' format, with the length of the text normalized to the width of the bar and each hit shown as a vertical line within the bar. This allows you to see the position where search results appear in the individual texts of a corpus.

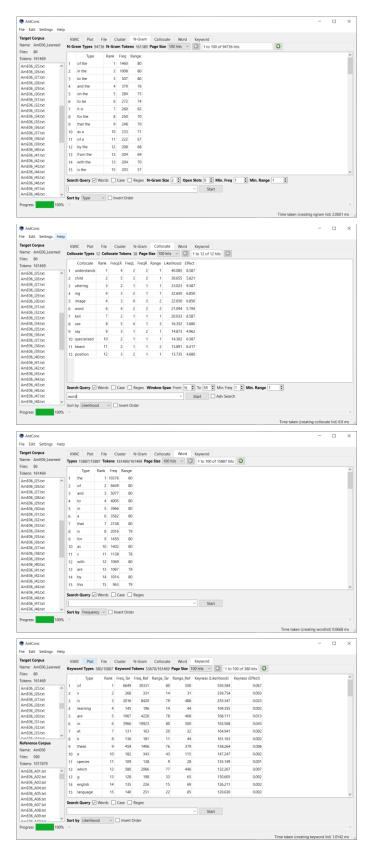


This tool shows the contents of individual texts. This allows you to investigate in more detail the results generated in other tools of AntConc.

Cluster Tool

The tool shows contiguous (together in a sequence) word patterns based on the search condition. This allows you to see common phrases that appear in the target texts.

AntConc ile Edit Settings Help						- 0	×
arget Corpus iame: AmE06_Learned				luster kens 4	N-Gram 2 Page Siz	Collocate Word Keyword 100 hits ~ C 1 to 30 of 30 hits C	
ies: 80 okens: 161469		Cluster	Rank	Freq	Range		^
imE06_J01.txt ^		word and	1	4	3		
imE06_J02.txt	2	word is	1	4	3		
imE06_J03.txt imE06_J04.txt	3	word the	3	3	2		
imE06_J05.txt	4	word for	- 4	2	1		
AmE06_J06.txt AmE06_J07.txt	5	word learning	4	2	1		
	6 1	word of	4	2	2		
AmE06_J08.txt AmE06_J09.txt	7	word or	4	2	2		
AmE06_J10.txt	8	word argument	8	1	1		
AmE06_J11.txt		word by	8	1	1		
AmE06_J12.txt AmE06_J13.txt	10	word consider	8	1	1		
mE06_J14.bxt		word hackman	8				
imE06_J15.txt		word if	8				
imE06_116.bxt		word in					
AmE06_J17.txt AmE06_J18.txt			8		1		
AmE06_J19.txt		word initial	8		1		
AmE06_J20.txt	15	word kali	8	1	1		~
AmE06_J21.txt AmE06_J22.txt	Sea	rch Query 🔽 Wo	irds 🗌	Case	Regex	Ister Size 2 C Min. Freq 1 C Min. Range 1	
AmE06 J23.txt	wo	ed.				V Start Adv Search	
AmE06_J24.txt v						rm Position On Left On Right On Left/Right	
rogress 100%		t by Frequency \	() L.) In	vert Ord	ter Search	trm Position On Left On Kight On Left/Kight	



N-Gram Tool

This tool scans the entire corpus for all 'N'-sized clusters (e.g., 2-word clusters, 3-word clusters, ...). This allows you to find common expressions in a corpus.

Collocate Tool

This tool shows words that appear frequently within a certain distance of the search term (i.e., collocates). This allows you to find which words co-occur with other words in a corpus.

Word List:

This tool counts all the words in the corpus and presents them in an ordered list. This allows you to find which words are the most frequent in a corpus.

Keyword List:

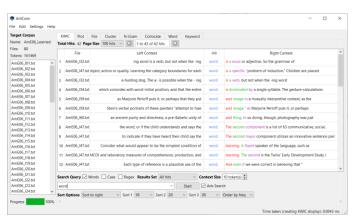
This tool shows words that appear unusually frequently in the target corpus in comparison with the words in the reference corpus based on a statistical measure (i.e., 'keywords'). These words can be considered to be characteristic of the target corpus. The settings can also be changed to show words that appears unusually infrequently in the target corpus compared with the reference corpus (i.e., 'negative keywords'.

KWIC (' Key-Word-In-Context') Tool

This tool shows search results in a concordance or 'KWIC' (Key-Word-In-Context) format. This allows you to see how words and phrases are commonly used in a corpus of texts.

The following steps produce a set of concordance lines from a corpus and demonstrate the main features of this tool.

 Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu. The files corpus

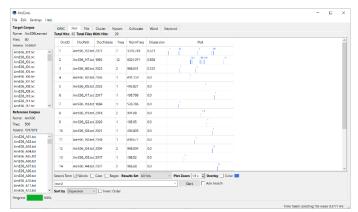


Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

- 2) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 3) Choose the size of the results set to be presented using the "Result Set" combobox widget.
- 4) Choose the number of words to be displayed on either side of the search term using the "Context Size" spinbox widget.
- 5) Click on the "Start" button to start the search and wait for the results to be displayed.
- 6) Use the "Sort options" to rearrange the concordance lines by row ID, file name, or the position of the word. The first widget allows you to quickly order the concordance lines by the words to right or left of the center word, or choose no ordering, or using a custom order. The next three widgets allow you to choose the order parameters: 1L, 2L... are words to the left of the target word, 'C' is the center word, and 1R, 2R... are words to the right of the center word. The final widget allows you to order the results by the frequency of the pattern determined by the sorting parameters (the "Order by freq" option) or alphabetically (the "Order by value" option). The default "Order by freq" option is strongly recommended as it will allow you to easily identify the most commonly occurring patterns in the target corpus. After adjusting the sort options, click on the "Start" button to regenerate the concordance lines.
- 7) The total number of concordance lines generated (Total Hits) is shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- 8) Double-clicking on any cell in the results window will cause the software to jump to the File tool (see the relevant section of this document) where you can view the hit exactly as it appears in context in the original file.
- 9) If you want to filter the results, select the desired rows and then press the "Delete" key to remove the selected rows or press "SHIFT+ Delete" to keep the selected rows removing all the others.
- 10) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Plot Tool

This tool shows concordance search results plotted in a 'barcode' format, with the length of the text normalized to the width of the bar and each hit shown as a vertical line within the bar. This allows you to see the position where search results appear in the individual texts of a corpus. An example of the use of the Plot Tool is in determining where specific content words appear in a technical paper, or where an actor or story character appears through a play or novel.



The following steps produce a set of plot results from a corpus and demonstrate the main features of this tool.

- 1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
- 2) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 3) Choose the size of the results set to be presented using the "Result Set" combobox widget.
- 4) Use the "Plot Zoom" widget to control the size of the plot and the degree of detail to be shown.
- 5) Click on the "Start" button to start the search and wait for the results to be displayed.
- 6) Use the "Sort by" option to rearrange the plots according to the various parameters shown.
- 7) The total number of hits and total number of plots are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- 8) Double-clicking on any cell in the results window will cause the software to jump to the File tool (see the relevant section of this document) where you can view the hit exactly as it appears in context in the original file.
- 9) By checking the "Overlay" option and choosing an appropriate color (by clicking on the color box), existing results can be overlaid with new results for different searches. This allows you to see how different search queries are related and/or overlap.
- 10) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

File Tool

This tool shows the text of individual files. This allows you to investigate in more detail the results generated in other tools of *AntConc*.

The following steps produce a view of the original file and demonstrate the main features of this tool.

 Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu.

AntConc	- o ×
le Edit Settings Help	
Target Corpus	KWIC Plot File Cluster N-Gram Collocate Word Keyword
lame: AmE06_Learned	File Hits 118 File Types 644 File Tokens 2022 File Name AmE06_J01.txt
Here: 80 Science: 101400 Amr60x, 2011bit Amr60x, 2011bit Amr60x, 2011bit Amr60x, 2011bit	Index of the second se
AmE06_J24.txt v	the Start Adv Search
Progress 100%	
	Time taken (creating file view) 0.0435 ms

The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

- 2) Double-click a file in the "Target Corpus" list on the left of the main window to view its contents. Alternatively, select a file in the "Target Corpus" list and click "Start" in the tool interface. The File tool will automatically be selected, and the contents of the file will be shown.
- 3) To highlight particular search query results in the display, enter a search query and click "Start". See the relevant section in this document explaining the "Words", "Case", and "Regex" search term options. Words in the file that match the query will be automatically highlighted.
- 4) Use the "Hit Location" widget to jump to different hits in the file. Alternatively, use the keyboard shortcut for your operating system (see the SHORTCUTS section).
- 5) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Cluster Tool

1)

The tool shows adjacent word groups based on the search condition. This allows you to see how words and phrases are commonly used in a corpus of texts. In some cases, this tool can be seen as summarizing the results generated in the KWIC tool

The following steps produce a set of clusters and demonstrate the main features of this tool.

Select a corpus using the "Corpus Manager"

available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

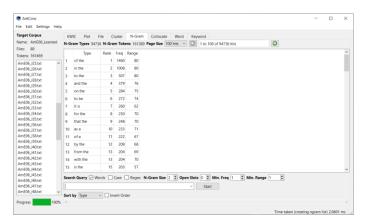
- 2) Enter a search guery in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- Choose the various parameters to filter the number of clusters to be shown: cluster size (number of 3) words in the cluster), minimum cluster frequency, and minimum cluster range (number of files)
- 4) Click the "Start" button to start the search and wait for the results to be displayed.
- 5) Use the "Sort by" option to rearrange the ordering of the results.
- Use the "Search Term Position" options to determine if the results will show clusters that start with the 6) search query terms ("On Left"), end with the search query terms ("On Right") or can either start or end with the search query terms ("On Left/Right").
- 7) The total number of cluster types ("Cluster Types") and combined total count of all the cluster tokens ("Cluster Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the 8) relevant section of this document) where you can view concordance lines for that cluster across the whole corpus.
- 9) Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

N-Gram Tool

This tool scans the entire corpus for all 'N'-sized clusters (e.g., 2-word clusters, 3-word clusters, ...). This allows you to find common expressions in a corpus.

The following steps produce a set of n-grams and demonstrate the main features of this tool.

1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the



"Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

Choose the various parameters to filter the number of n-grams to be shown: n-gram size (number of 2) words), open slots (number of slots in the n-gram that can take multiple values), minimum n-gram frequency, and minimum n-gram range (number of files).

- 3) Click on the "Start" button to start the search and wait for the results to be displayed. If a search query is entered, only n-grams that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 4) Use the "Sort by" option to rearrange the ordering of the results.
- 5) The total number of n-gram types ("N-Gram types") and combined total count of all the n-gram tokens ("N-Gram Tokens") are shown at the top of the tool window.
 When no hits are found, a warning will be shown on the screen.
- 6) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that n-gram across the whole corpus.
- 7) For entries that contain open slots, Shift + Double-click on the "Type" entry in the results window to show the variants that can fit in the open slots via the "Open Slot Viewer".
- 8) Several menu preferences are available with this tool. (See the relevant sections in this document for explanations).
- 9)

Collocate Tool

This tool shows words that appear frequently within a certain distance of the search term (i.e., collocates). This allows you to find which words cooccur with other words in a corpus.

The following steps produce a set of collocates and demonstrate the main features of this tool.

 Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the

En	ntry: word	l + a			
Г	S1	S1 freq			
1	is	3			
2	learning	1			

arget Corpus ame: AmE06.Learned		WIC Plot	File	Cluste		-Gram	Colle			yword	
es: 80	Col	locate Types	12 Coll	ocate To	kens 3	3 Page	Size 10	0 hits ~	0 1	12 of 12 hits 🕤	
uens: 161469		Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect		
E06_J25.txt	1	understands	1	- 4	2	2	1	40.085	8.587		
E06_J26.txt	2	child	2	5	3	2	1	30.655	5.821		
E06_J27.txt E06_J28.txt	3	uttering	3	2	1	1	1	23.023	9.587		
06_129.bit	4	ing	4	3	2	1	1	22.650	6.850		
606_J30.txt	5	image	4	3	0	3	2	22.650	6.850		
06_J31.brt 06_J32.brt	6	word	6	4	2	2	2	21.094	5.194		
06_J32.bit	7	kali	7	2	1	1	1	20.033	8.587		
06_J34.txt	8	use	8	5	- 4	1	3	16.352	3.680		
06_J35.txt	9	say	9	3	1	2	1	14.873	4.962		
06_J36.brt 06_J37.brt	10	specialized	10	2	1	1	1	14.362	6.587		
506_J38.txt	11	heard	11	2	1	1	2	13.891	6.417		
06_J39.txt	12	position	12	3	2	1	1	13.735	4.680		
06_140.txt 06_141.txt											
E06_J42.txt											
E06_J43.txt											
06_J44.txt 06_J45.txt					_						
E06_146.bxt	Sea	rch Query 🗹	Words	Cas	e 🗌 R	egex N	Vindow	Span From	5L C	o SR 🗘 Min. Freq 1 🗘 Min. Range 1 🗘	
506 J47.txt	wo							~	s	t Adv Search	

"Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".

- 2) Choose the various parameters to filter the types of collocates to be shown: window span (possible positions left and right of the search query terms, where the collocate can appear), minimum collocate frequency, and minimum collocate range (number of files).
- 3) Enter a search query in the search box. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 4) Click on the "Start" button to start the search and wait for the results to be displayed.
- 5) Use the "Sort by" option to rearrange the ordering of the results.
- 6) The total number of collocate types ("Collocate Types") and combined total count of all the collocate tokens ("Collocate Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- 7) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that collocate across the whole corpus.

 Advanced searches are available with this tool. Several menu preferences are also available with this tool. (See the relevant sections in this document for explanations).

Word Tool

This tool counts all the words in the corpus and presents them in an ordered list. This allows you to find which words are the most frequent in a corpus.

The following steps produce a word list and demonstrate the main features of this tool.

Target Corpus	KV	/IC Plot	Eile	Clus	ter N-	word	
Name: AmE06_Learned						to 100 of 15887 hits	
Files: 80	ijpe	 1500//13 	500/ 100	uens 10	1403/1014	10 100 01 1360/ HIS	
Tokens: 161469		Type	Rank	Freq	Range		
	1	the	1	10376	80		
AmE06_J25.txt	2	of	2	6649	80		
AmE06_J27.txt	3	and	3	5077	80		
AmE06_J28.txt	4	to		4005	80		
AmE06_J29.txt		in		3966	80		
AmE06_J30.txt AmE06_J31.txt	5						
AmE06_J32.txt	6	a		3562	80		
AmE06_J33.txt	7	that	7	2138	80		
AmE06_J34.txt	8	is	8	2016	79		
AmE06_J35.txt AmE06_J36.txt	9	for	9	1450	80		
AmE06_J36.btt AmE06_J37.btt	10	as	10	1402	80		
AmE06_J38.bxt	11	s	11	1138	78		
AmE06_J39.txt	12	with	12	1069	80		
AmE06_J40.txt	13	are	13		78		
AmE06_J41.txt AmE06_J42.txt			14	1014	80		
AmE06_J43.txt	14	by					
AmE06_J44.txt	15	this	15	963	79		
AmE06_J45.txt	Sear	ch Query	Words		se 🗌 Re		
AmE06_J46.txt AmE06_J47.txt	1					Start	
	v				rt Order	Start	

- 1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
- 2) Click on the "Start" button to start the processing and wait for the results to be displayed. If a search query is entered, only words that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 3) Use the "Sort by" option to rearrange the ordering of the results.
- 4) The total number word types ("Word types") and combined total count of all the word tokens ("Word Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- 5) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that word across the whole corpus.
- 6) Several menu preferences are available with this tool. (See the relevant sections in this document for explanations).

Keyword Tool

This tool shows words that appear unusually frequently in the target corpus in comparison with the words in the reference corpus based on a statistical measure (i.e., 'keywords'). These words can be considered to be characteristic of the target corpus. The settings can also be changed to show words that appears unusually infrequently in the target corpus compared with the reference corpus (i.e., 'negative keywords'.

rget Corpus me: AmE06_Learned		VIC Plot word Types 38				Collocate 670/161469	Keyword 100 hits v G 1 to	100 of 380 hits	
rs: 80 kens: 161469		Туре	Rank				 Keyness (Likelihood)	Keyness (Effect)	
mE06_J25.txt		of	1	6649	30331	80	550.584	0.067	
nE06_J26.bxt nE06_J27.bxt	2	x	2	268	331	14	339.754	0.003	
nE06_J28.txt	3	is	3	2016	8420	79	255.347	0.023	
nE06_J29.bxt	4	learning are	4	145	196 4226	14	169.355	0.002	
AmE06_J30.txt AmE06_J31.txt AmE06_J32.txt		are in	6	3966	4226	78	165.568	0.013	
	6	et .	7	3900	19923	20	164,941	0.043	
nE06_J33.txt	8	et k	8	131	103	20	164,941	0.002	
ference Corpus	9	k these	9	459	1406	76	158,264	0.002	
me: AmE06	10	e	10	459	343	43	130.204	0.008	
s: 500	10	e species	11	102	138	43	147.247	0.002	
ens: 1017879		which	12	580	2066	77	132.267	0.007	
E06_A01.txt	13	9	13	128	198	33	130.605	0.002	
nE06_A03.txt	14	9 english	14	135	226	15	126,211	0.002	
nE06_A04.txt nE06_A05.txt	15	language	15	140	251	22	120.630	0.002	
nE06_A03.bit nE06_A06.bit nE06_A07.bit nE06_A08.bit nE06_A09.bit		ch Query 🗹 \					 ✓ Start 		

The following steps produce a keyword list and demonstrate the main features of this tool.

- 1) Select a corpus using the "Corpus Manager" available from the File menu. Alternatively, create a quick corpus by choosing the "Create Quick Corpus" from the file menu. The files contained in the corpus are shown in the left frame of the main window under "Target Corpus".
- 2) Click on the "Start" button to start the processing and wait for the results to be displayed. If a search query is entered, only words that match the query will be shown. See the 'SEARCH OPTIONS' section in this document for an explanation of the "Words", "Case", and "Regex" search term options.
- 3) Use the "Sort by" option to rearrange the ordering of the results.

- 4) The total number keyword types ("Keyword types") and combined total count of all the word tokens ("Keyword Tokens") are shown at the top of the tool window. When no hits are found, a warning will be shown on the screen.
- 5) Double-click on any cell in the results window to cause the software to jump to the KWIC tool (see the relevant section of this document) where you can view concordance lines for that keyword across the whole corpus.
- 6) Several menu preferences are available with this tool. (See the relevant sections in this document for explanations).

SEARCH OPTIONS

Search queries can be composed of full words or word fragments, with or without wildcards. The basic syntax roughly follows the Common Elementary Query Language (CEQL). See https://cwb.sourceforge.io/ceql.php for more details). Searches can be either "case insensitive" (default) or "case sensitive" by activating or deactivating the "Case" search term option. Searches can also be made using full regular expressions by activating the "Regex" option. With "regex" option, each word-level regular expression needs to be separated by whitespace. To make regex expressions case-aware, select the "Case" option. For details on how to use regular expressions, consult one of the many texts on the subject, e.g., Mastering Regular Expressions (O'Reilly & Associates Press) or type "regular expressions" in a web search engine to find many sites on the subject (e.g., http://www.regular-expressions.info/quickstart.html). AntConc supports Perl regular expressions including Unicode character classes, e.g., \p{Letter}, even though the software is built using Python.

ADVANCED SEARCHES

By clicking on the "Advanced Search" button (available in most tools), more complex searches become possible.

- The "Search Query List" option allows you to import a set of search queries. You can do this in one of three ways: 1) Type each individual search query in the entry box and click "Add"; 2) Drag and drop the list of search queries into the viewer below the entry box; 3) Copy and paste the list of search queries into the viewer below the entry box; 3) Copy and paste the list of queries, each line will be treated as a separate search query. This feature allows you to use a large set of search queries without having to retype them each time. Any search query accepted in the main interface can be used.
- The "Context Search List" option allows you to define search queries that must match within a certain context window around the main search term(s). For example, to search for "student" or "students" appearing at least three words to the left or right of the word "university," add "university" to the "Search Query List", and then add "student" and "students" to the "Context Search List" list. Finally, set the "Context Search List" "Window Span" as "From 3L" and "To 3R".
- The "SQL Search List" option allows you to adjust query conditions by applying conditions directly on tables in the database that stores the corpus through a series of 'join' operations. For example, the following entry in the list will join a custom "genres" table in the database (created through the Corpus Manager) with the main "corpus" table, applying a condition that the "genre" column entry for the file must be marked as "academic". The two tables are joined via the common "doc_id" column:

["genres", "genre = 'academic'", "doc_id"]

The format for the list entry should be a JSON array, with three components, "table" + "condition" + "join column"

MENU OPTIONS

Menu options are divided into three groups, "File", "Edit", "Settings" and "Help". The options available in each group will be described below.

- Create Quick Corpus...
 - This option is for quickly creating a temporary corpus. You will be asked to choose the files you
 want to add to your corpus. Then, the software will create a "temp " corpus using the default
 settings of the Corpus Manager. Any existing "temp" corpus will be overwritten.
- Open Corpus Manager...
 - This option opens the Corpus Manager, where you can choose prebuilt corpora from the default library, add or delete corpora from a user library, or create custom corpora from raw files. See the Corpus Manager for complete details.
- Clear Tool/All Tools/All Tools and Files
 - These options will reset the interface
- Save Current Tab Results...
 - This option allows results to be exported in a file format. Direct copying and pasting of results from the interface are also possible
- Save Current Tab Datatables...
 - This option allows complete tables of results from the corpus database to be exported to a set of .csv files. All relevant information about results can be found in these files.
- Import Settings From File.../Export Settings To File...
 - These options allow the state of the software to be saved and reloaded at a different time
- Restore Default Settings
 - This option resets the state of the software to when it was first launched. All custom settings are lost.

<EDIT>

- Select All
 - This option selects all results in the results window. The same effect can be achieved using the standard keyboard shortcut for "Select All". See the SHORTCUTS section for more details.
- Сору
 - This option selects any text in the results window. The same effect can be achieved using the standard keyboard shortcut for "Copy". See the SHORTCUTS section for more details.

<SETTINGS>

- Global Settings (applied to all tools in the interface)
 - Colors decides the main highlight color (e.g., for highlighted words in the File tool)
 - Files decides how the paths to files are shown
 - Fonts decides the font family, size, and style of the font for the main interface
 - Searches lists all wildcard available in the system (note that these cannot be edited)
 - Statistics decides how values are displayed in the interface
 - Tags decides how tags are displayed in the interface
 - Language Direction decides how to display results (especially for the KWIC concordance tool) depending on the language direction. For example, choose the default "Left-to-right" option for language such as English. Choose the "Right-to-left" option for languages such as Arabic.
- Tool Settings
 - o KWIC
 - Display Options decides the colors used to highlight the sort order
 - Other Options
 - Choose to show or hide the file names in the display
 - Choose to show or hide the search term in the display. This option is useful for allowing instructors to quiz students on possible words to fit the gap.

- View Style decides which view to use (table/graphic or graphic)
- Display Options decides how results are displayed
- Statistics decides the parameters for determining the dispersion measure
- Other Options decides various parameters for sizing/displaying the plot graphs
- o Cluster
 - Display Options decides what information is shown in the results window
 - Filter Options decides if clusters can only span cross whitespace boundaries or can include other characters (e.g., punctuation)
- o N-Gram
 - Display Options decides what information is shown in the results window
 - Filter Options decides if clusters can only span cross whitespace boundaries or can include other characters (e.g., punctuation)
- \circ Collocate
 - Likelihood Measure + Threshold
 - Choose the statistic and cut-off point (threshold) for inclusion of words in the collocates list. Words below the cut-off-point are deemed to appear frequently together with the query term by chance.
 - Effect Size Measure
 - Choose the statistic used determine the strength of relationship between the query term and collocate and a cut-off point (threshold) for inclusion of words that meet the minimum effect size.
- $\circ \quad \text{Word}$
 - List Type decides what type of list is shown. The "Word List" option will show the frequency of words in the corpus. The "Headword List" option will conflate the frequencies of individual words in the corpus under a grouping headword (if available). The family members for each headword will be displayed with their frequencies of occurrence appearing in parentheses.
 - Display Options decides what information is shown in the results window
- \circ Keywords
 - Display Options decides what information is shown in the results window
 - Negative Keywords decides to show words in the target corpus that appear unusually infrequently in the target corpus compared with the target corpus.
 - Likelihood Measure + Threshold decides the statistic and cut-off point for inclusion of words in the keyword list. Words below this cut-off-point are deemed to appear frequently in the target corpus compared with the reference corpus by chance.
 - Effect Size Measure + Threshold decides the measure used determine the strength keyness and a cut-off point for inclusion of words that meet the minimum effect size.

[Appropriate effect size measures are still being debated in the field, so the default setting is to show all values for this measure. With the default settings, keywords are ranked according to their likelihood measure scores. This equates to ranking keywords according to p-values, which raises several questions/problems. However, it is the current standard in the field and results tend to show that ranking by likelihood leads to more intrinsically intuitive results than those generated when an effect size measure is used. The current selection of likelihood measures and effect size measures are inspired by the work of Andrew Hardie of Lancaster University.]

< HELP >

- View Help
 - This option shows the current help guide as a PDF file.
- Show License

- This option shows the license agreement that you agree to when using the software.
- Show Version History
 - This option shows the complete history of releases, detailing new features, bug fixes, and major updates.
- About AntConc
 - This option shows the release version, release date, copyright information, and acknowledgments for the software.

Corpus Manager

Choosing/Saving a pre-built corpus	Building/Saving a corpus from raw files	Building/Saving a corpus from a word list
Mildler topology X	Riddine topoManya X	Mildler topology X
Verden Verdender Frankrise Verdender Verd	Verlander Constante	Variance Marchine
tony driving (51)	Come Tools	
Supported any Reverse Gappa	Suppled and and foreign support	Supplicated angle of Revenue Garges Received Receive

The Corpus Manager is a multi-purpose tool used to load and save pre-built corpora, create and save a new corpus from raw files, or create and save a new corpus from a word list. The three different scenarios are explained below.

Choosing/Saving a pre-built corpus

Choose the "Pre-built" corpus option.

- If you click this option, a list of pre-built corpora available in the "Default Corpus Library" of AntConc is shown in the top-right of the window. Choose one of these by selecting it and then clicking "Choose" or double clicking on the list item. A description of the chosen corpus will be shown in the top-left of the window.
- 2. If you have previously saved pre-built corpora into the "User Corpus Library" of AntConc they will be shown in the bottom-right of the

et Corpus Ref	ference Corpus	
rpus Source		Pre-Built Corpora
Pre-built 🔿 Ra	w files O Word list	Default Corpus Library
Active Corpus		AmE06.db
AmE06 J Jearned	db Load Save Clear	AmE06_A_press_reportage.db
Description		AmE06_B_press_editorial.db AmE06_C_press_reviews.db
Cateopry	Description	AmE06_D_religion.db
full_name	American English 2006: Learned	AmE06_E_skills_and_hobbies.db
short name	AmE06 Learned	AmE06_F_popular_lore.db AmE06_G belles lettres.db
file.count	80	AmE06_H_misc.db
token count	161469	AmE06_J_learned.db
type_count	17804	Choose
language	English	Encore.
date	2006	User Corpus Library
mode	witten	Add File(s) Add Directory Delete
license	Free for research purposes	
reference	Potts, A., & Baker, P. (2012). Does semantic tagging identify cultural change in British and American English? International Journal of Corpus Linguistics. 17(3), 295-324.	
summary	Natural Sciences, Medicine, Mathematics, Social and Behavioral Sciences, Political Science, Law, Education, Humanities, Technology and Engineering.	
encoding	utf-8-sig	
token_definition	[\p(L])+	Choose
	- · · · · · · · · · · · · · · · · · · ·	Choose

window. You can add other pre-built corpora to the "User Corpus Library" list using the "Add File(s)" or "Add Directory" options. Once you have added the files, you can choose one of these by selecting it and then clicking "Choose" or double clicking on the list item. Any corpora built from raw files or a word list (see below) will also be shown here. A description of the chosen corpus will be shown in the top-left of the window.

3. Pre-built corpora can be loaded directly by clicking on the "Load" button on the top-left of the window and choosing the corpus file. Pre-built corpora can be saved to a new location by clicking on the "Save" button and saving the file.

Click the "Return to Main Window" button at the bottom of the Corpus Manager to return to the main windows and start using the corpus. The corpus files will be shown in the left pane of the main window.

Building/Saving a corpus from raw files

Choose the "Raw Files" corpus option. Several 'builder options' will appear on the top-right of the window. Follow the steps below:

- 1. Choose the files to be included in your corpus
 - Use the "Add File(s)" or "Add Directory" options to choose your raw files. You can choose plain TEXT (.txt), WORD (.docx) and PDF (.pdf) files.
- Choose one or more of the build options as necessary (OPTIONAL)
 - a. Choose a name for your custom corpus. A default name is provided.

orpus Source				Raw Files Corpus Builder	r			
) Pre-built 🖲 R	w files O Word list			Corpus files				
Active Corpus				Add File(s) Add D	lirectory Clear			No. of files 6
alice_corpus.db		Save	Clear	alice_in_wonderland.c	docx			^
Description				alice_in_wonderland.p				
Cateogry full_name	Descripti alice_corpus	on		alice_in_wonderland.t alice_through_the_loc alice_through_the_loc	oking_glass.docx oking_glass.pdf			
short_name	alice_corpus			alice through the log	kinn class bit			Ŷ
file_count	6			Options				
token_count	783			Name	alice_corpus			
type_count	150			Indexer	simple_word_in	dexer		~
encoding	utf_8_sig			Encoding	UTF-8 O C	ther Western	Europe (latin_1)	
token_definition	[\p(L])+			Token Definition		Show Toker	Definition Settin	ias
ignore_header	False				Add File(s)		Clear	No. of files (
ignore_items	False				Hud Hie(s)	nou orrectory	Crear	NO. OT THES Y
number_replace	False			Metadata Tables				
format	raw_files							
indexer_type	word							
indexer	simple_word_indexer			Headword/Grouping I	Add File		Preview	Clear
				Create				100

- b. Decide the indexer used to process the raw files.
 - i. For simple files with no annotation or part-of-speech (POS) tagging, the "simple_word_indexer" will work well.
 - ii. For simple files that have been part-of-speech (POS) tagged, the "simple_word_pos_headword_indexer" will work well.
 - iii. For simple files that have been part-of-speech (POS) tagged using the Biber Tagger, the " simple_word_bibertag_indexer " will work well
 - iv. For other files with more complex structures, different indexers will become available over time.
- c. Choose the character encoding of the files
 - i. The default option (UTF-8) is the standard in the field. Many other encodings are also available in the "Other" options.
 - If you are unsure of what encoding to use, it is recommended to initially choose the default UTF-8 option. Later, if you find that your files appear corrupted in the various tool displays, it probably means that the encoding is wrong, and you should determine the correct encoding to use. For Word (.docx) and PDF (.pdf) files, the default encoding should generally be fine.
 - ii. For more information on the Unicode standards see:
 - http://www.cs.tut.fi/~jkorpela/unicode/guide.html
 - http://www.unicode.org/
 - http://www.unicode.org/Public/5.0.0/ucd/UCD.html
 - http://www.unicode.org/Public/UNIDATA/PropList.txt
 - http://www.unicode.org/charts/
- d. Decide is the definition of a token (word) in the corpus.
 - i. In some cases you may only want to include tokens (words) comprised of the letters a-zA-Z in your corpus, whereas other times, you might want to include tokens comprised of letters, numbers, apostrophes etc. The token definition determines what tokens your corpus is comprised of.
 - ii. If you click on the "Show Token Definition Settings" button, a new window will open where you can choose your definition. AntConc offers three ways to choose a token definition: "Character Classes", "User-Defined Characters", and "User-Defined Regex":
 - "Character Classes": This is the default option and is the most comprehensive. Click the various options to add characters to the definition. These classes are fully Unicode compliant, meaning that they can handle data in any language, including all European languages and Asian languages. For example, the default option "Letters" refers to 'letters' in the broadest sense, including all English letters (a to z, A to Z) and also all Japanese and Chinese 'letter' characters.

- 2. "User-Defined Characters": This is a simple option whereby all characters you type in the text edit box will be included in the final token definition.
- 3. "User-Defined Regex": This is another simple option whereby all characters that match the given regex (regular expression) will be included in the final token definition.
- iii. If you active the "Ignore header" option, you can choose a starting and ending tag for your corpus files headers. Any text between these two tags will be ignored when the corpus is created.
- iv. If you active the "Ignore non-embedded tags" option, you can choose a starting and ending tag for elements in your corpus files that you want to ignore. Any text between these two tags will be ignored when the corpus is created.
- v. If you active the "Ignore embedded tags" option, you can choose a tag maker for token elements in your corpus files that you want to ignore. Any text appearing after the tag marker will be ignored when the corpus is created.
- vi. Using the "Token Testing Area", you can test your token definition by typing or copy/pasting a text into the left-hand text box, clicking the "Test" button, and checking in the right-hand text box to see what tokens will be generated.
- vii. After defining your corpus token definition, click "Apply"
- e. Choose metadata tables (if available)
 - i. If you click on "Add File(s)" or "Add Directory", you can choose optional metadata tables that will be stored as SQLite database tables together with your raw corpus data. The information in these metadata tables must be aligned with the column names used in the existing tables of the corpus. To understand the default table structure, open the corpus database in an SQLite database reader (e.g., https://sqlitebrowser.org/) and view the different tables.
 - ii. Once the corpus is built, you will be able to form search queries on the main corpus using values in these tables as conditional elements.
- f. Choose a headword/grouping list (if available)
 - i. If you click on "Add File", you can choose an optional headword/grouping list that will be used to map words to headwords or grouping categories. Existing headwords (e.g., those generated by a POS tagger) will be overwritten by these headwords.
 - ii. Once the corpus is built, you will be able to form search queries on the main corpus using these headword/grouping terms as conditional elements.
- 3. Create the corpus
 - a. To complete the corpus building process, click the "Create" button.
 - b. Once the corpus is created, a basic description will be displayed in the left of the window.
 - c. The corpus will be available in the "User Corpus Library" and can be saved to another location by clicking on "Save" button next to the Active Corpus label.

Building/Saving a corpus from a word list

Choose the "Word List" corpus option. Several 'builder options' will appear on the top-right of the window. Follow the steps below:

- Under "Corpus information file", click the "Add File" button to choose a "Corpus information file" in CSV or TSV format that matches the "corpus_info" table structure used by AntConc.
- 2. Under "Word list information file", click the "Add File" button to choose a "Word list information file" in CSV or TSV format that

rpus Source Pre-built O Ri Active Corpus my_bioscience_w Description	aw files @ Word list ordfist.db Save Clear	Word List Corpus Builder Corpus Information File Add File Clear Z/projects/AntConc/test, files/plos_one_pdfs/Mord_yeau/ts/corpus_info.cov
Cateogry full_name	Description my_bioscience	Word list information file Add File Clear
short_name	my_bioscience	Z/projects/AntConc/test files/plos one pdfs/Word results/wordlist info.csv
encoding	utf_8_sig	
token_definition	[\p(L)]+	Word list data file
indexer_type	word	Add File Clear
format	word_list	Z/projects/AntConc/test_files/plos_one_pdfs/Word_results/wordlist.csv
type_count	9525	Name
token_count	74042	my_bioscience_wordlist
file_count	10	my_oroscience_woronsi
		Create 100%

matches the "wordlist_info" table structure used by AntConc.

3. Under "Word list data file", click the "Add File" button to choose a "Word list data file" in CSV or TSV format that matches the "wordlist" table structure used by AntConc.

[To understand the various formats required, load one of the existing pre-built corpora, generate a word list, and then save the database tables via the File menu "Save Current Tab Datatables" option. In the saved .zip file, you will find examples of the table formats that you need to use. Alternatively, open the corpus database in an SQLite database reader (e.g., <u>https://sqlitebrowser.org/</u>) and view the different tables. An SQLite database reader should allow you to export the tables as a CSV/TSV files that you can use as templates for your own wordlist corpus creation.]

- 4. Create the corpus
 - a. To complete the corpus building process, click the "Create" button.
 - b. Once the corpus is created, a basic description will be displayed in the top-left of the window.
 - c. The corpus will be available in the "User Corpus Library" and can be saved to another location by clicking on "Save" button next to the Active Corpus label.

[Note that a wordlist corpus can only be used as a reference corpus for use when generating keyword lists. Using a wordlist-based corpus in other tools will result in an error.]

SHORTCUTS

Here is a list of useful shortcuts (including some of the useful standard shortcuts on the operating system).

- CTRL/COMMAND + TAB: Toggles clockwise through the different tools in the tab bar.
- ALT + Tool Number: Selects a specific tool (e.g., ALT+1 for the KWIC Tool, ALT+2 for the Plot tool).
- SHIFT + CTRL/COMMAND + TAB: Toggles anti-clockwise through the different tools in the tab bar.
- CTRL/COMMAND + C: Copies the currently selected text.
- CTRL/COMMAND + A: Selects all text in the window.
- F4 (Win): Reveals the complete list of options in a 'combobox' widget (e.g., the search history in the search query box).
- ARROW KEYS: For any 'combobox' widgets (e.g., the KWIC search query box) or 'spinbox' widgets (e.g., the KWIC context size), the 'UP' and 'DOWN' arrow keys on the keyboard can be used to change the value of the option.
- CTRL/COMMAND + O: Opens the "Corpus Manager".
- CTRL/COMMAND + G: Opens the "Global Settings".
- CTRL/COMMAND + T: Opens the "Tool Settings".
- CTRL/COMMAND + H: Toggles the view "(H)ide" of "Show" setting for the KWIC tool file name column.
- CTRL/COMMAND + '+': Zooms in the "Plot Tool" display.
- CTRL/COMMAND + '-': Zooms out the "Plot Tool" display.
- CTRL/COMMAND + F: Searches for the next hit in the "File Tool" display.
- SHIFT + CTRL/COMMAND + F: Searches for the previous hit in the "File Tool" display.

NOTES

- If you have any suggestions for improving the software or notice any bugs, please post them in the AntConc Discussion Group (https://groups.google.com/g/antconc). Indeed, many of the improvements and updates made to the software have been due to the comments of users around the world, for which I am very grateful. The AntConc Discussion Group is also a good place to discuss how you are using the software and any challenges that your face.
- If you find the software useful in your research, teaching, or learning, you may consider making a small donation to support the future development of this tool. A link to the donation page can be found here: https://www.laurenceanthony.net/software/antconc/

- You may also be interested in becoming an AntConc patron. Depending on the level of support, this option will give you priority support with direct access to the developer (Laurence Anthony), and various other benefits. A link to the donation page can be found here
- <u>https://www.patreon.com/antlab</u>

CITING/REFERENCING ANTCONC

Use the following method to cite and reference AntConc according to the APA style guide:

Anthony, L. (YEAR OF RELEASE). AntConc (Version VERSION NUMBER) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

For example, if you download AntConc 4.0.0, which was released in 2021, you will cite/reference it as follows: Anthony, L. (2021). AntConc (Version 4.0.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

ACKNOWLEDGEMENTS

I would like to say thank you to the users of *AntConc* who have taken the trouble to post feedback on the software and make suggestions for improvements and/or changes. A very special thank you goes to all those who have very generously supported the project either through single donations via <u>PayPal</u> or becoming a <u>Patreon supporter</u>. A complete list of individual acknowledgments can be found in the Help menu - "About AntConc" menu.

The development of *AntConc* has been supported by a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (C): No. 23501115, a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (B): No. 18700658, a Japan Society for Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (B): No. 16700573, and a WASEDA University Grant for Special Research Projects: No. 2004B-861.

KNOWN ISSUES

None at present.