

# 워드스미스 툴 6.0 특강

A Special Seminar on WordSmith Tools 6.0

Dr. Mike Scott

University of Aston and Lexical Analysis Software Ltd.  
mike.scott@aston.ac.uk  
www.lexically.net

일 시: 2014. 8. 20(수) 16:00 ~ 17:30

장 소: 부산대학교 인문관 컴퓨터실(401호)

주 최: 코퍼스언어학연구회

(Research Association of Corpus Linguistics)

주 관: 부산대학교 글로벌 영상번역 창의인재 교육-연구 혁신사업단  
(BK플러스)



Pusan National University

# 워드스미스 툴 6.0 특강

## A Special Seminar on WordSmith Tools 6.0

Dr. Mike Scott

University of Aston and Lexical Analysis Software Ltd.

mike.scott@aston.ac.uk

www.lexically.net

일 시: 2014. 8. 20(수) 16:00 ~ 17:30

장 소: 부산대학교 인문관 컴퓨터실(401호)

주 최: 코퍼스언어학연구회

(Research Association of Corpus Linguistics)

주 관: 부산대학교 글로벌 영상번역 창의인재 교육-연구 혁신사업단

(BK플러스)



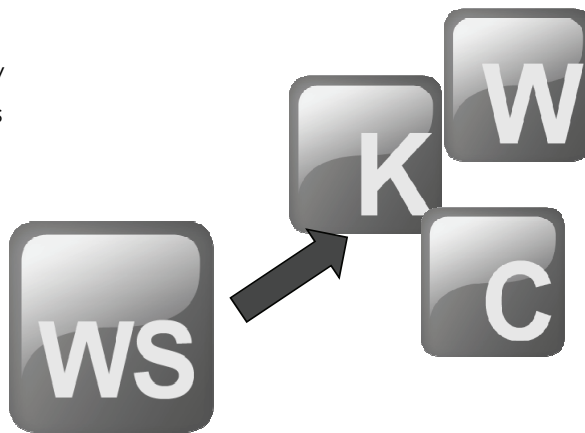
Pusan National University

# Contents

- Developing WordSmith Tools ..... 1
- How to Use WordSmith Tools: ..... 5  
Install WS, Wordlist, Keywords, Concord,  
Index, Clusters, Keyclusters

## Developing WordSmith Tools

- Mike Scott
- Aston University
- Lexical Analysis Software Ltd



1

For WordSmith you may if you wish use the single user licence for the individual PCs, but it may be easier to use a network licence:

Here is the network registration code.

Registered to: Korea Maritime and Ocean University  
Other details: temporary  
Registration code: SD00.0617.2331.6428.1570.6688.A8AU

WordSmith version: 06  
For site use, not individual use  
(<http://lexically.net/downloads/version6/HTML/index.html?networkdefaults.htm>)  
see  
[http://www.lexically.net/wordsmith/version6/faqs/network\\_installation.htm](http://www.lexically.net/wordsmith/version6/faqs/network_installation.htm) for detailed instructions on installation.

Issued: 13 August 2014  
This registration OK until: 14 September 2014  
Download from <http://lexically.net/wordsmith/version6/>

2



## The aim

- personal use at home
- many languages
- assortment of tools
- lexical focus
- download only

3



## History

- *MicroConcord* 1993 (with Tim Johns)
- *WordSmith Tools* 1.0 1996, Oxford University Press
  - 2.0, 1997 OUP
  - 3.0, 1999 OUP
- 4.0, 2004 OUP (Unicode)
  - 5.0, 2008 Lexical Analysis Software (WSConcgram)
  - 6.0, 2012 Lexical Analysis Software (clouds, scripts, phrase frames, colour categories)

4



## Learning

- Key words function developed almost by accident
- Many settings & features arose at user request

5



## Problems

- developing an intuitive interface
- providing help
- ignoring language-specific knowledge
- technology constantly changing

6



## Demonstration and hands-on

- See [http://www.lexically.net/wordsmith/support/get\\_started\\_guides.html](http://www.lexically.net/wordsmith/support/get_started_guides.html) if you need help...
- a word list
- a concordance from the word list
- viewing the source text
- a key word list
- exporting results to Word or Excel

7



## Help

- [http://www.lexically.net/wordsmith/support/get\\_started\\_guides.html](http://www.lexically.net/wordsmith/support/get_started_guides.html)
- (no Guide yet in Korean)
- <http://www.lexically.net/downloads/version6/HTML/index.html>
- free licence:

**Registered to: Korea Maritime and Ocean University**

**Other details: temporary**

**Registration code: SA00.0613.2777.1799.1570.8653.A8A3**

This registration is OK until: 14 September 2014

Download from <http://lexically.net/wordsmith/version6/>

When registering (Utilities | Registration), please PASTE these details in EXACTLY as you see them here.

8

## How to Use WordSmith Tools:

Install WS, Wordlist, Keywords, Concord,  
Index, Clusters, Keyclusters

August 20, 2014

Se-Eun Jhang & Sungmin Lee  
(Korea Maritime and Ocean University)

Busan National University

1

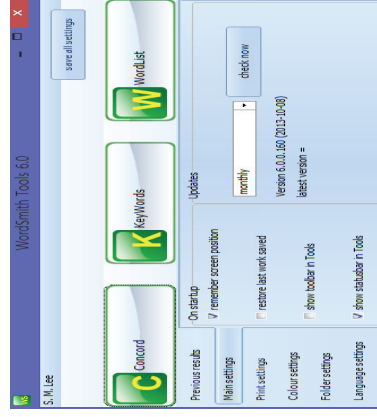
## Contents

- How to install WordSmith
- How to get Wordlist
- How to get Keywords
- How to get Concordance
- How to get Index
- How to get Clusters
- How to get Keyclusters

2

## Advantages of Wordsmith Tools

- 편리한 화면 구성
- 상세한 매뉴얼
- 다양한 통계제공



3

## www.lexically.net/wordsmith



4



# Purchase WS

WordSmith Tools

HOME :: VERSION 5.6 :: DOWNLOAD

purchase

UPGRADE

Finding word patterns. PC software published by Oxford University Press since 1996

Concord ... for finding all instances of a word or phrase.  
▶ [read more](#)

Keywords ... helps find salient words in a text or set of texts.  
▶ [read more](#)

5

# Purchase WS

LEXICAL ANALYSIS SOFTWARE LTD  
supplying and supporting WordSmith Tools

Your payment is handled by Digital River, who provide a secure site which handles credit/debit cards, orders, invoices, PayPal, WebMoney, DEAL, etc.

You receive your licence and registration code by email in seconds.

[How to order & payment procedures](#)  
[Universities: Ordering the Purchase Order](#)  
[Upgrades Policy](#)

**New purchases:**

- Single user Licence (for one person, stand-alone on your laptop or PC)  
▶ **Single user licence** (single licence bundles)
- Network Licence (for multiple users, individual users)  
▶ Up to 10 users site licence
- Up to 50 Users Licence (installed on a network and available to up to 10)  
▶ Up to 10 users site licence
- Up to 50 Users Licence (installed on a network and available to up to 50)  
▶ Up to 50 users site licence

6

# Single User License

Secure Site

Secure order process  
[Click here for security info]

언어: 한국어 | 가격 표시 통화: 대한민국 원

귀하의 장바구니

구매 제품 이름	비중	단기	수량	가격	적기
1. WordSmith Tools Single User (206310)	정식 버전	KRW 95,039 / USD 90.07	1	KRW 95,039 / USD 90.07	<input type="checkbox"/>

쿠폰 코드를 입력하십시오(쿠폰이 있는 경우):

구매하기 ▶ 총액: KRW 95,039 / USD 90.07

7

# Single User License

Customer information | Additional information | Review and submit | Finished

Language: English

You can indicate a separate delivery or billing address, if needed, at a later point in time.

**Personal Information**

License to:  
 first name, last name  
 company name

Solution  
First Name \*  
Last Name \*  
Street Address \*  
Additional Address Information  
ZIP / Postal Code \*  
City \*  
State / Province  
Country \*  
VAT ID  
Phone  
FAX

Payment Option

Payment type \*  
Billing currency \*  
New Credit Code  
Please enter your coupon code (if available)

Note: fields marked with an asterisk (\*) are required fields.

Next

8

# License Key

Research Association of Corpus Linguistics [.] .txt - 메모장  
 WordSmith Tools 5.0  
 Purchased: 21/01/2012  
 By: jhang@hu.ac.kr  
 Type: single user (student discount)

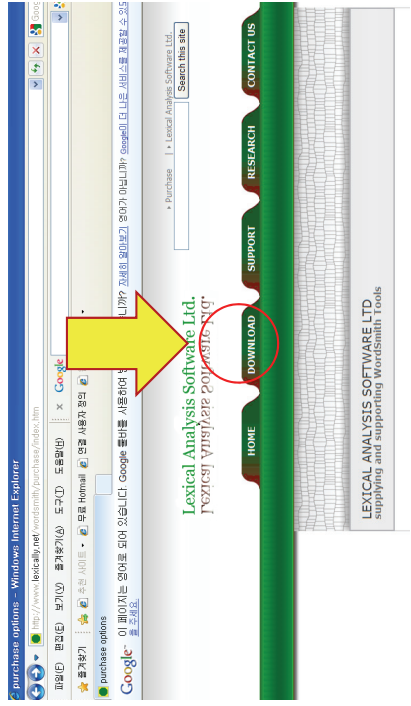
\*\*\*\*\*  
 PLEASE DON'T JUST EMAIL THE WHOLE LIST TO EVERYONE\*\*\*\*\*  
 PRINT AND CUT THE LINES INTO SEPARATE SLIPS FOR DISTRIBUTION\*\*\*\*\*  
 \*\*\*\*\*



REGISTRATION Number 6 of 10  
 Registered to: S. M. Lee  
 Other details: Research Association of Corpus Linguistics  
 Registration code: SE40.0789.6770.17xx.xxxx

Paste these in EXACTLY as given here without adding anything

# Installation of WS



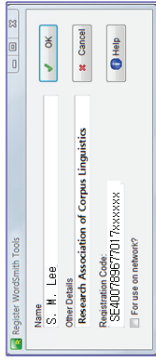
# Installation of WS 6.0



만약 계속하지 못하는  
 경우가 생기면 Cancel  
 나중에 계속 설치하고  
 자 하면 바탕화면에  
 생성된 WS setup 클  
 릿하면 좌측화면이 또  
 나옴



# Installation of WS



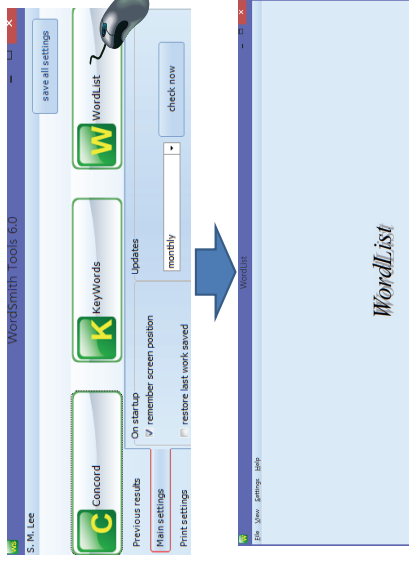
# Wordlist

- This program generates word lists based on one or more plain text files.
- Word lists are shown both in alphabetical and frequency order.
- They can be saved for later use, edited, printed, copied to your word-processor, or saved as text files.

출처 : WordSmith 6.0 manual

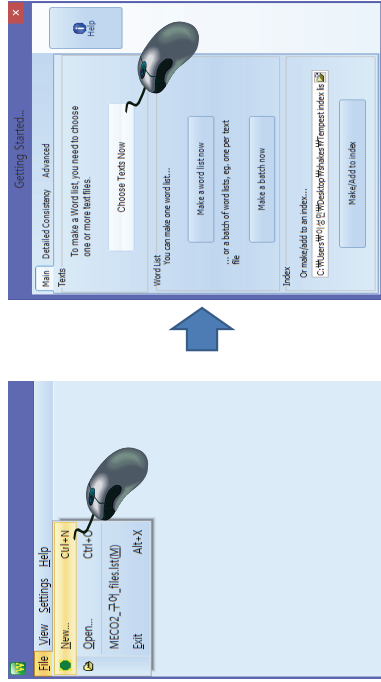
13

# WordList



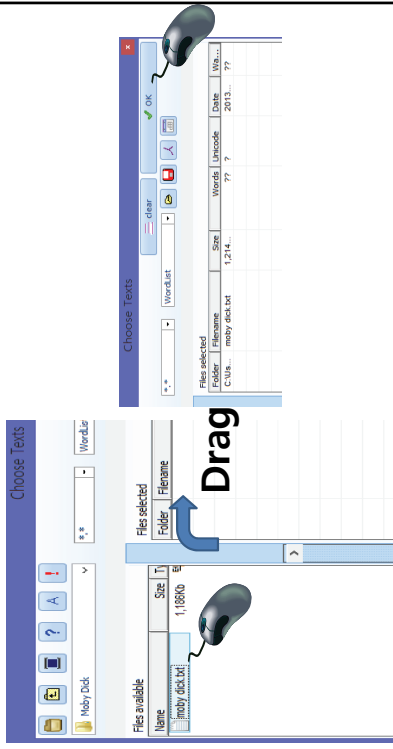
14

# Start WordList



15

# Upload Txt files



16

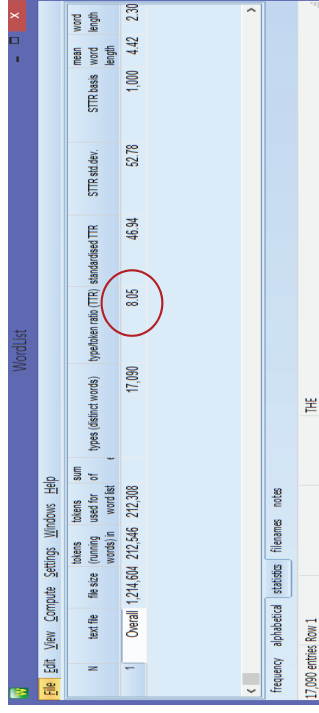
# Click "Make a wordlist now"



# Save WordList



# Check Statistics for Moby Dick



# TTR (Type-Token Ratio)

- $TTR = (\text{number of type} / \text{number of token}) \times 100$
- TTR means the ratio of type and token. The way to get the TTR is the number of types over the number of tokens multiplied by 100.
- For example, the TTR of texts used by native speakers is 22.67. On the other hand, the TTR of texts by non-native speakers is 13.48. This shows that native speakers use a greater variety of vocabulary than non-native speakers.

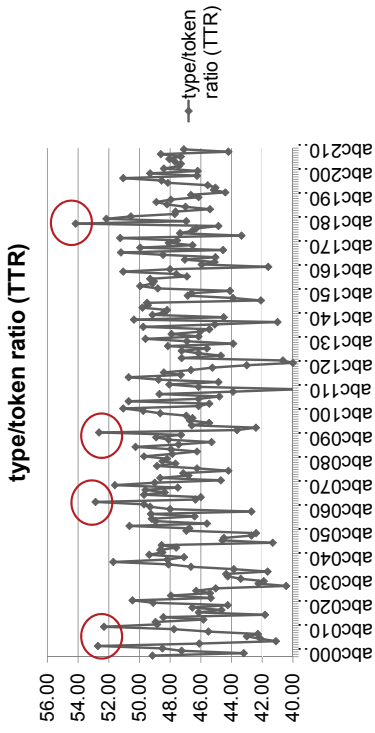
## Python Coding For Customized TTR Analysis

```

split-the-sentence.py - c:\python22\split-the-sentence.py
File Edit Format Run Options Windows Help
inputfile = "abc01000.txt", format=(int(LI/size))
count = 0
for i in range(1, len(inputfile)):
    count = count + 1
    if re.search('\.\s+', w):
        print words[0:count]
        count = 0
        words = []
        file=open('abc01000.txt', 'r')
        f = file.readlines()
        file.close()
def main(argv):
    # default#
    inputfile = 'The Tempest.txt'
    size = 500
    # program name
    if len(argv) > 1:
        inputfile = argv[1]
    if argv:
        try:
            opt, args = getopt.getopt(argv, "h:is:", ["help", "size="])
        except getopt.GetoptError:
            print "split-the-sentence.py -i <inputfile> -s <word size>, file=
            argv[0]
            sys.exit(0)
        for opt, arg in opt:
            if opt == "-h":
                print ("Usage: %s <inputfile> -i <inputfile> -s <word size> or,
                print ("Usage: %s <inputfile> -i <inputfile> -s <word size> or,
                sys.exit(0)
            elif opt in ("-i", "--inputfile"):
                inputfile = arg
            elif opt in ("-s", "--size"):
                size = int(arg)
    split_files(inputfile, size)
if __name__ == '__main__':
    main(sys.argv)

```

## Moby Dick TTR ratio per 100 words



## Save a Wordlist Files



## Keyness

The term "key word", though it is in common use, is not defined in Linguistics. This program identifies key words on a mechanical basis by comparing patterns of frequency. (A human being, on the other hand, may choose a phrase or a superordinate as a key word.) A word is said to be "key" if

- it occurs in the text at least as many times as the user has specified as a Minimum Frequency
- its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a *p* value specified by the user.

### positive and negative keyness

A word which is *positively* key occurs *more* often than would be expected by chance in comparison with the reference corpus.  
A word which is *negatively* key occurs *less* often than would be expected by chance in comparison with the reference corpus.

출처 : WordSmith 6.0 manual

# How Are Keywords Calculated?

**Definition of Keyness**

The term 'key word' though it is common use, is not adequate. A human being, on the other hand, may choose a phrase or a word to use in a particular context. This is done on the basis of the frequency in the text when compared with its frequency in other texts.

**positive and negative keyness**

A word which is negatively key occurs less often than would be expected.

**typical key words**

These are words which occur in a higher proportion in one text than in another. First, there will be proper nouns. Proper nouns are often key words. This can be avoided by specifying a higher Minimum Frequency.

Second, there are key words that human beings use of things and places. This program does not guess synonyms, and so will miss key words unless you are comparing word-lists based on word lists.

Third, there are key words which are 'key' because of their frequency in a particular text. These are 'key' words. This program does not guess synonyms, and so will miss key words unless you are comparing word-lists based on word lists.

See also: [New Key Words Calculated](#), [Definition of Key Words](#)

**How Key Words are Calculated**

The 'key words' are calculated by comparing the frequency of each word in the text with its frequency in the other text. The program does this by comparing the frequency of each word in the text with its frequency in the other text. The program does this by comparing the frequency of each word in the text with its frequency in the other text.

# Log-likelihood Calculator

<http://ucrell.lancs.ac.uk/lwizard.html>

To use this wizard, type in frequencies for one word and the corpus sizes and press the calculate button.

Notes:

- Please enter plain numbers without commas (or other non-numeric characters) as they will confuse the calculator!
- The LL wizard shows a plus or minus symbol before the log-likelihood value to indicate overuse or underuse respectively in corpus 2.
- The log-likelihood value itself is always a positive number. However, my script compares relative frequencies between the two corpora 1 relative to corpus 2.

# How to Calculate Log Likelihood

Log likelihood is calculated by constructing a contingency table as follows:

	Corpus 1	Corpus 2	Total
Frequency of word	a	b	a+b
Frequency of other words	c-a	d-b	c-d-a-b
Total	c	d	c+d

Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two. We need to calculate the expected values (E) according to the following formula:

$$E_i = \frac{M_i \cdot C_i}{N}$$

$$-2 \ln \Lambda = 2 \sum_{i=1}^k O_i \ln \left( \frac{O_i}{E_i} \right)$$

In our case  $M_1 = c$ , and  $M_2 = d$ . So, for this word  $E_1 = c \cdot (a+b) / (c+d)$  and  $E_2 = d \cdot (a+b) / (c+d)$ . The calculation for the expected values (E), whereas we need to calculate the expected values (E) according to the following formula. We can then calculate the log-likelihood value according to the formula above.

The higher the G2 value, the more significant is the difference between two frequency scores. For these tables, a G2 of 3.84 or higher is significant at the level of  $p < 0.05$  and a G2 of 6.63 or higher is significant at  $p < 0.01$ .

- 95th percentile, 5% level,  $p < 0.05$ ; critical value = 3.84
- 99th percentile, 1% level,  $p < 0.01$ ; critical value = 6.63
- 99.9th percentile, 0.1% level,  $p < 0.001$ ; critical value = 10.83
- 99.99th percentile, 0.01% level,  $p < 0.0001$ ; critical value = 15.13

# Getting Keywords

WordSmith Tools 6.0

KeyWords

KeyWords

Keyword: 4:30centuries.lws

Exit

# Moby Dick vs Contem19C

Click Target Corpus and Reference Corpus  
Click Make a keyword list now



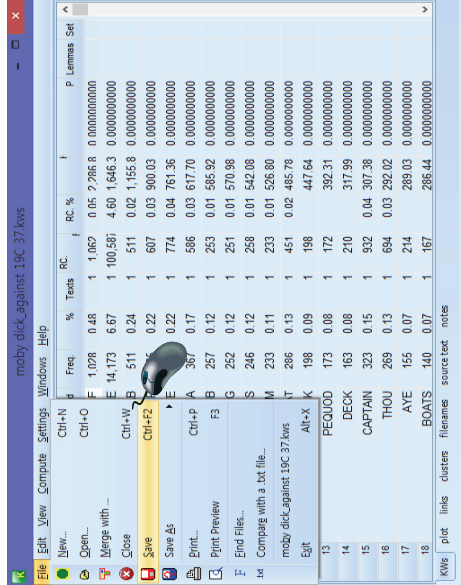
# Positive Keywords

N	Key word	Freq.	%	Tests	RC	f	RC %	P	Lemmas Set
1	WHALE	1,028	0.48	1	1,062	0.05	2,286.8	0.0000000000	
2	THE	14,173	6.67	1	100,567	4.60	1,646.3	0.0000000000	
3	AHAB	511	0.24	1	511	0.02	1,155.8	0.0000000000	
4	SHIP	465	0.22	1	607	0.03	900.03	0.0000000000	
5	YE	466	0.22	1	774	0.04	761.36	0.0000000000	
6	SEA	367	0.17	1	566	0.03	617.70	0.0000000000	
7	STUBB	257	0.12	1	253	0.01	585.92	0.0000000000	
8	QUEEJEG	252	0.12	1	251	0.01	570.98	0.0000000000	
9	WHALES	246	0.12	1	288	0.01	512.08	0.0000000000	
10	SPERM	233	0.11	1	233	0.01	526.80	0.0000000000	
11	BOAT	266	0.13	1	461	0.02	465.78	0.0000000000	
12	STARBUCK	198	0.09	1	198	0.01	447.64	0.0000000000	
13	PEQUOD	173	0.08	1	172	0.00	392.31	0.0000000000	
14	DECK	163	0.08	1	210	0.03	317.99	0.0000000000	
15	CAPTAIN	323	0.15	1	932	0.04	307.38	0.0000000000	

# Negative Keywords

N	Key word	Freq.	%	Tests	RC	f	RC %	P	Lemmas Set
631	HOUSE	34	0.02	1	2,124	0.10	-239.35	0.0000000000	
632	MY	587	0.28	1	10,789	0.49	-224.96	0.0000000000	
633	HERSELF	7	0.00	1	1,732	0.08	-264.24	0.0000000000	
634	HE	1,876	0.89	1	27,912	1.26	-270.09	0.0000000000	
635	T	287	0.14	1	7,283	0.33	-301.92	0.0000000000	
636	LADY	8	0.00	1	1,582	0.09	-302.45	0.0000000000	
637	MOTHER	17	0.01	1	2,340	0.11	-315.23	0.0000000000	
638	TO	4,511	2.12	1	28,575	1.31	-524.08	0.0000000000	
639	WAS	1,627	0.77	1	62,071	2.84	-397.65	0.0000000000	
640	YOU	884	0.42	1	18,514	0.85	-536.37	0.0000000000	
641	S-AD	300	0.14	1	10,385	0.47	-465.53	0.0000000000	
642	MRS	13	0.01	1	4,166	0.19	-460.64	0.0000000000	
643	I	2,114	1.00	1	39,279	1.80	-483.50	0.0000000000	
644	HAD	767	0.36	1	19,895	0.91	-487.77	0.0000000000	
645	MR	61	0.03	1	7,146	0.33	-519.72	0.0000000000	
646	SHE	330	0.16	1	24,549	1.12	-2,664.4	0.0000000000	
647	SHE	114	0.05	1	19,612	0.91	-2,838.1	0.0000000000	

# How to Save as Keyword List

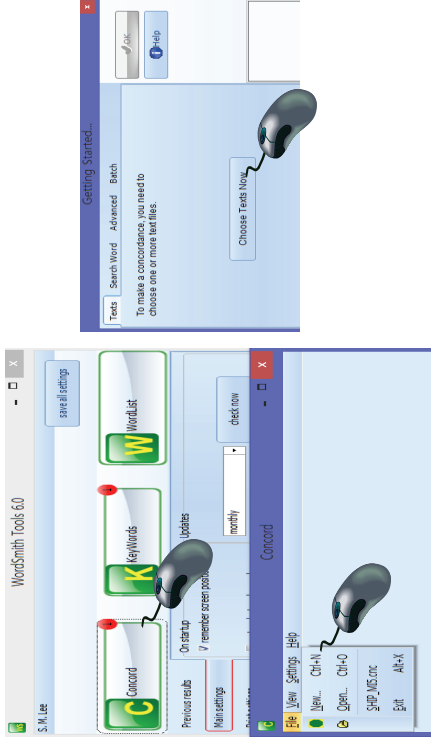


# Concordance

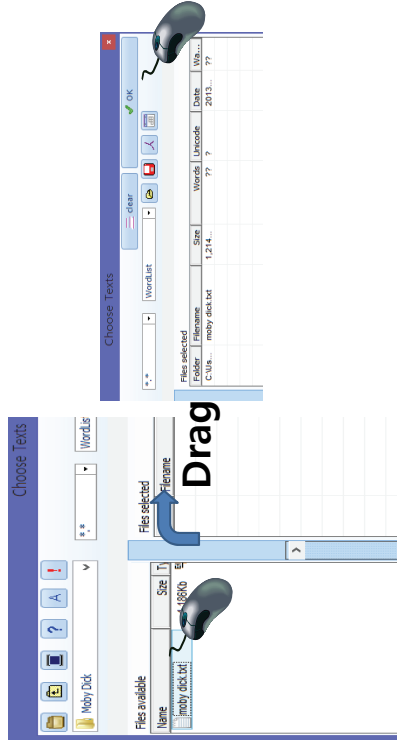
- To use it you will specify a search word, which Concord will seek in all the text files you have chosen.
- It will then present a concordance display, and give you access to information about collocates of the search word.

출처 : Wordsmith 6.0 manual

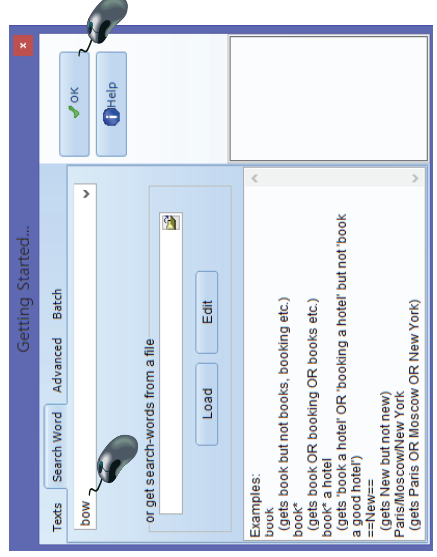
# Concordance



# Upload Txt file



# Insert Search Word





# Save Concordance

File Edit View Compute Settings Windows Help

Concordance

1 would make the eye of that whale a bow-window some five feet long. Ah,  
 2 you sail then—Haloah! stop! make a bow before you go—Avast heaving  
 3 on the gny, embattled, **bartering** bow, but only to drag dark Ahab after  
 4 the horizontal piece in the **boat's** bow for bracing the knee against in  
 5 to be necessarily tantalizing his **boat's** bow with its tail—these allusions of his  
 6 carelessly reclining in his own **boat's** bow. He was a darkly-tanned, burly,  
 7 is thus annihilated; often the **boat's** bow is knocked off, or the thigh-board,  
 8 colling some spare line in the **boat's** bow, "did you never hear that the ship  
 9 mate, was standing up in his **boat's** bow, and with all the reckless energy  
 10 it projected beyond his **whale-boat's** bow; but the sea that had stove its  
 11 and lay there reversed; **broken** bow to shattered stem. At last he  
 12 for the stranger. Drawing across **her** bow, he perceived that in accordance  
 13 round the boat upon its axis, **its** bow, by anticipation, was made to  
 14 when only some twenty of her **naked** bow-ribs are inserted, and the keel is  
 15 your corrupt companion. Oh! do **not** bow and look surprised; you know the  
 16 and hammered against the **opposing** bow. "Drive, drive in your nails, oh ye  
 17 in the surface again for off the **other** bow but within a few weeks of Ahab's

concordance collocates plot patterns clusters timeline filenames source text notes

0% 0%

# How to get Collocations

File Edit View Compute Settings Windows Help

Concordance

1 would make the eye of that whale a bow-window some five feet long. Ah,  
 2 you sail then—Haloah! stop! make a bow before you go—Avast heaving  
 3 on the gny, embattled, **bartering** bow, but only to drag dark Ahab after  
 4 the horizontal piece in the **boat's** bow for bracing the knee against in  
 5 to be necessarily tantalizing his **boat's** bow with its tail—these allusions of his  
 6 carelessly reclining in his own **boat's** bow. He was a darkly-tanned, burly,  
 7 is thus annihilated; often the **boat's** bow is knocked off, or the thigh-board,  
 8 colling some spare line in the **boat's** bow, "did you never hear that the ship  
 9 mate, was standing up in his **boat's** bow, and with all the reckless energy  
 10 it projected beyond his **whale-boat's** bow; but the sea that had stove its  
 11 and lay there reversed; **broken** bow to shattered stem. At last he  
 12 for the stranger. Drawing across **her** bow, he perceived that in accordance  
 13 round the boat upon its axis, **its** bow, by anticipation, was made to  
 14 when only some twenty of her **naked** bow-ribs are inserted, and the keel is  
 15 your corrupt companion. Oh! do **not** bow and look surprised; you know the  
 16 and hammered against the **opposing** bow. "Drive, drive in your nails, oh ye  
 17 in the surface again for off the **other** bow but within a few weeks of Ahab's

concordance collocates plot patterns clusters timeline filenames source text notes

0% 0%

# Click Compute - Relationships

File Edit View Compute Settings Windows Help

Concordance

1 would make the eye of that whale a bow-window some five feet long. Ah,  
 2 you sail then—Haloah! stop! make a bow before you go—Avast heaving  
 3 on the gny, embattled, **bartering** bow, but only to drag dark Ahab after  
 4 the horizontal piece in the **boat's** bow for bracing the knee against in  
 5 to be necessarily tantalizing his **boat's** bow with its tail—these allusions of his  
 6 carelessly reclining in his own **boat's** bow. He was a darkly-tanned, burly,  
 7 is thus annihilated; often the **boat's** bow is knocked off, or the thigh-board,  
 8 colling some spare line in the **boat's** bow, "did you never hear that the ship  
 9 mate, was standing up in his **boat's** bow, and with all the reckless energy  
 10 it projected beyond his **whale-boat's** bow; but the sea that had stove its  
 11 and lay there reversed; **broken** bow to shattered stem. At last he  
 12 for the stranger. Drawing across **her** bow, he perceived that in accordance  
 13 round the boat upon its axis, **its** bow, by anticipation, was made to  
 14 when only some twenty of her **naked** bow-ribs are inserted, and the keel is  
 15 your corrupt companion. Oh! do **not** bow and look surprised; you know the  
 16 and hammered against the **opposing** bow. "Drive, drive in your nails, oh ye  
 17 in the surface again for off the **other** bow but within a few weeks of Ahab's

concordance collocates plot patterns clusters timeline filenames source text notes

0% 0%

# Upload Wordlist and Click Statistics

File Edit View Compute Settings Windows Help

Concordance

1 would make the eye of that whale a bow-window some five feet long. Ah,  
 2 you sail then—Haloah! stop! make a bow before you go—Avast heaving  
 3 on the gny, embattled, **bartering** bow, but only to drag dark Ahab after  
 4 the horizontal piece in the **boat's** bow for bracing the knee against in  
 5 to be necessarily tantalizing his **boat's** bow with its tail—these allusions of his  
 6 carelessly reclining in his own **boat's** bow. He was a darkly-tanned, burly,  
 7 is thus annihilated; often the **boat's** bow is knocked off, or the thigh-board,  
 8 colling some spare line in the **boat's** bow, "did you never hear that the ship  
 9 mate, was standing up in his **boat's** bow, and with all the reckless energy  
 10 it projected beyond his **whale-boat's** bow; but the sea that had stove its  
 11 and lay there reversed; **broken** bow to shattered stem. At last he  
 12 for the stranger. Drawing across **her** bow, he perceived that in accordance  
 13 round the boat upon its axis, **its** bow, by anticipation, was made to  
 14 when only some twenty of her **naked** bow-ribs are inserted, and the keel is  
 15 your corrupt companion. Oh! do **not** bow and look surprised; you know the  
 16 and hammered against the **opposing** bow. "Drive, drive in your nails, oh ye  
 17 in the surface again for off the **other** bow but within a few weeks of Ahab's

concordance collocates plot patterns clusters timeline filenames source text notes

0% 0%

Confirm Filename

The word list should be of the same corpus you have just concordanced.

C:\Users\1058\Desktop\1058\_1058\_2013\_2014\_102014\_1058\_1058\_1058

Case sensitive  
 Full lemmas processing  
 relation statistics  
 Specific Mutual Information

column for relation: Total

Help Cancel OK

# Kinds of Collocations

N	Word	With	Relation	Total	Total	L1	L2	L3	L4	L5	L6
1	BOW	bow	12.627	1	43	0	0	0	0	0	0
2	STARBOARD	bow	10.361	1	4	4	0	1	1	0	3
3	POINTS	bow	9.748	1	2	2	0	0	2	0	0
4	PERCEIVED	bow	9.641	1	2	0	0	2	0	0	0
5	WEATHER	bow	9.494	1	2	0	0	2	0	0	0
6	DESCRIBED	bow	8.989	1	2	0	0	2	0	0	0
7	ACROSS	bow	8.494	1	2	0	0	2	0	0	0
8	AGAINST	bow	8.494	1	2	0	0	2	0	0	0
9	HAND	bow	6.978	1	3	2	1	0	0	0	1
10	OFF	bow	6.326	1	3	2	1	0	0	0	1
11	BOAT	bow	6.289	1	4	1	3	0	0	0	1
12	FAR	bow	6.154	1	2	0	2	0	0	0	0
13	BEFORE	bow	5.829	1	3	1	3	0	0	0	0
14	OVER	bow	5.815	1	4	4	0	0	0	0	0
15	HER	bow	5.667	1	3	2	1	0	0	0	0
16	THREE	bow	5.560	1	2	2	0	0	0	0	0
17	ITS	bow	5.498	1	3	2	1	0	0	0	0

# Sort Relation values

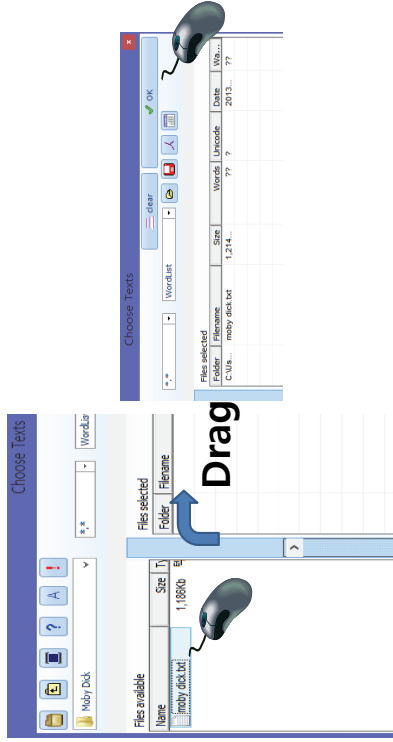
N	Word	With	Relation	Total	Total	f	L5
1	BOW	bow	12.627	1	43	0	0
2	STARBOARD	bow	10.361	1	4	4	0
3	POINTS	bow	9.748	1	2	0	0
4	PERCEIVED	bow	9.641	1	2	0	2
5	WEATHER	bow	9.494	1	4	4	0
6	DESCRIBED	bow	8.989	1	2	1	1
7	ACROSS	bow	8.494	1	2	2	0
8	AGAINST	bow	8.494	1	3	2	1
9	HAND	bow	6.978	1	3	2	1
10	OFF	bow	6.326	1	3	2	1
11	BOAT	bow	6.289	1	4	1	3
12	FAR	bow	6.154	1	2	0	0
13	BEFORE	bow	5.829	1	3	1	2
14	OVER	bow	5.815	1	4	4	0
15	HER	bow	5.667	1	3	2	1
16	THREE	bow	5.560	1	2	2	0
17	ITS	bow	5.498	1	3	2	1

# How to Create Index, Clusters, Keyclusters

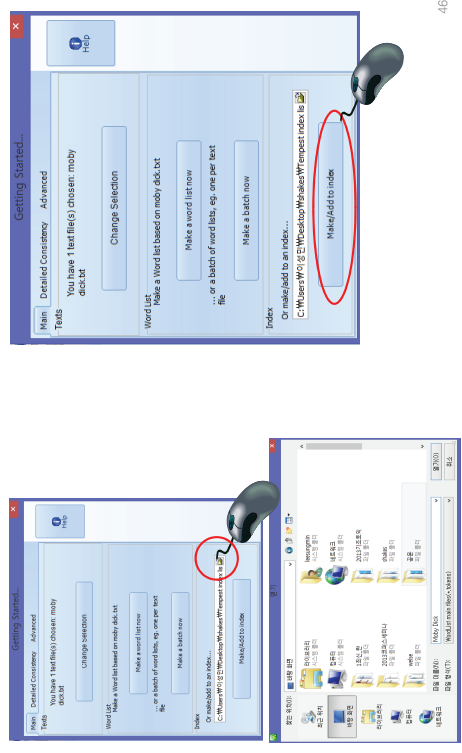
# INDEX: Make a Wordlist tool index

Wordlist  
Main Detailed Consistency Advanced  
You have 1 text file(s) chosen: moby.docx.txt  
Change Selection  
Wordlist: Make a Word list based on moby.docx.txt  
Make a word list now  
... or a batch of word lists - eg. one per text file  
Make a batch now  
Index  
Or make/add to an index...  
C:\Users\W101\g\l\W\Desktop\Wordlist\Wordlist index is [ ]  
Make/Add to index

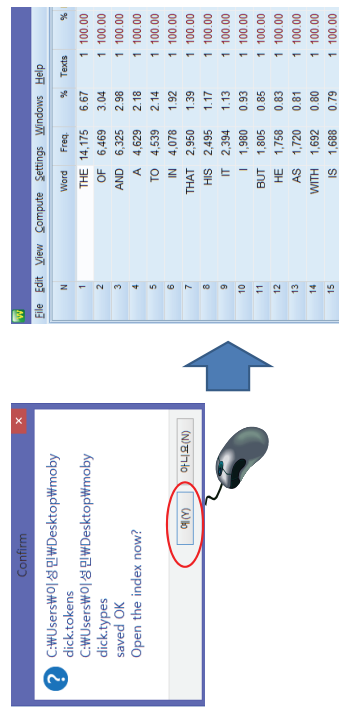
# Upload Txt file



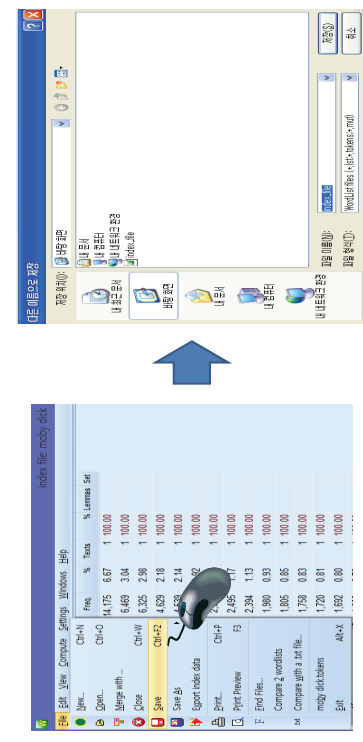
# Save/Click Make/Add to Index



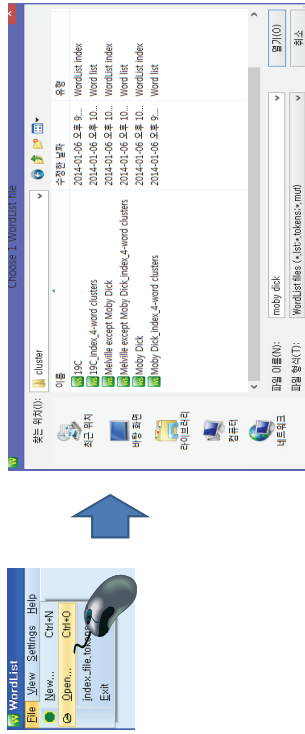
# Create Index File



# Save index files in your folder

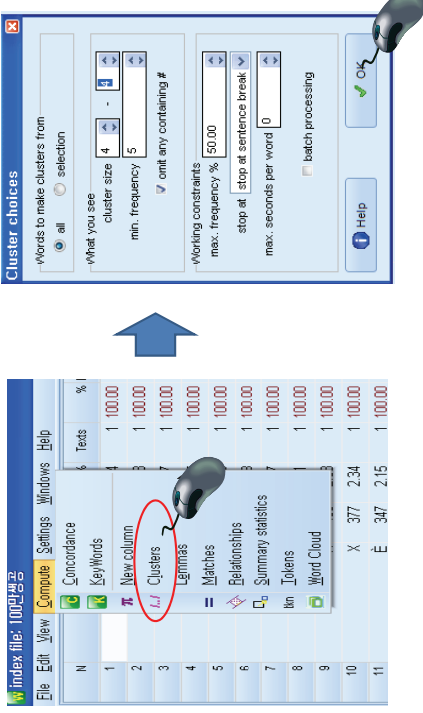


## Open saved index files



49

## Go compute & click Cluster choices



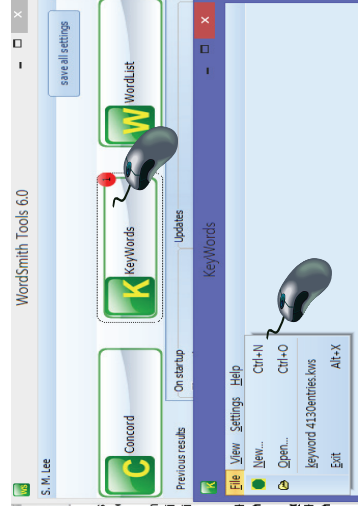
50

## Output of clusters

N	Word	Freq.	%	Texts	%
1	OF THE SPERM WHALE	42	0.02	1	100.00
2	THE SPERM WHALE S	27	0.01	1	100.00
3	AT THE SAME TIME	20		1	100.00
4	OF THE WHALE S	18		1	100.00
5	THE BOTTOM OF THE	17		1	100.00
6	THE OLD MAN S	16		1	100.00
7	IN THE ACT OF	12		1	100.00
8	THE WHITE WHALE S	12		1	100.00
9	AND AT THE SAME	11		1	100.00
10	AS IF IT WERE	11		1	100.00
11	I DON T KNOW	11		1	100.00
12	IN THE CASE OF	11		1	100.00
13	OF THE WHITE WHALE	11		1	100.00
14	SPERM WHALE S HEAD	11		1	100.00
15	BE IT SAID THAT	10		1	100.00

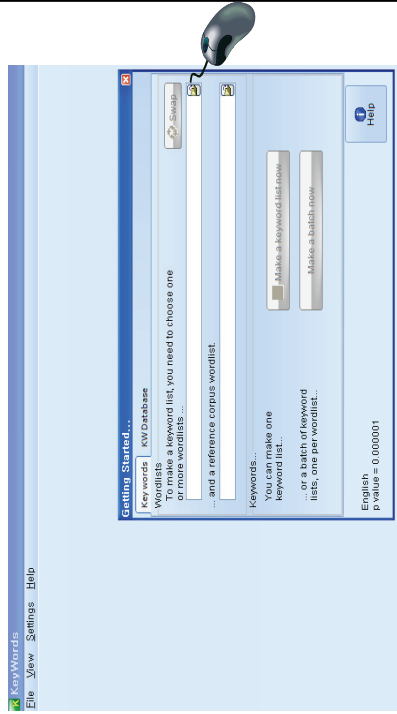
51

## Click a Keywords tool for making a key clusters list



52

# Start making Keyclusters



53

# Upload clusters list



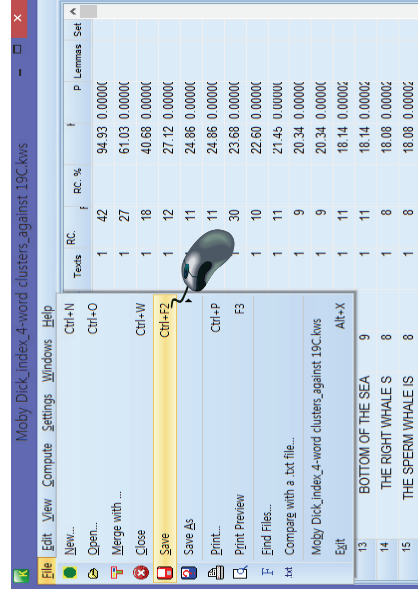
54

# Output of Plus Keyclusters

N	Keyword	Freq	%	RC	f	RC %	P	Lemmas	Set
1	OF THE SPERM WHALE	42	0.02	1	42	94.93	0.00000		
2	THE SPERM WHALES	27	0.01	1	27	61.03	0.00000		
3	OF THE WHALES	18		1	18	40.68	0.00000		
4	THE WHITE WHALES	12		1	12	27.12	0.00000		
5	SPERM WHALE S HEAD	11		1	11	24.86	0.00000		
6	OF THE WHITE WHALE	11		1	11	24.86	0.00000		
7	THE OLD MAN S	16		1	30	23.68	0.00000		
8	THE GREAT SPERM WHALE	10		1	10	22.60	0.00000		
9	BE IT SAID THAT	10		1	11	21.45	0.00000		
10	ON BOARD THE PEQUOD	9		1	9	20.34	0.00000		
11	OF THE PEQUOD S	9		1	9	20.34	0.00000		
12	OF THE SEA AND	9		1	11	18.14	0.00000		
13	BOTTOM OF THE SEA	9		1	11	18.14	0.00000		
14	THE RIGHT WHALE S	8		1	8	18.08	0.00000		
15	THE SPERM WHALE IS	8		1	8	18.08	0.00000		

55

# How to Save Keyclusters List



56