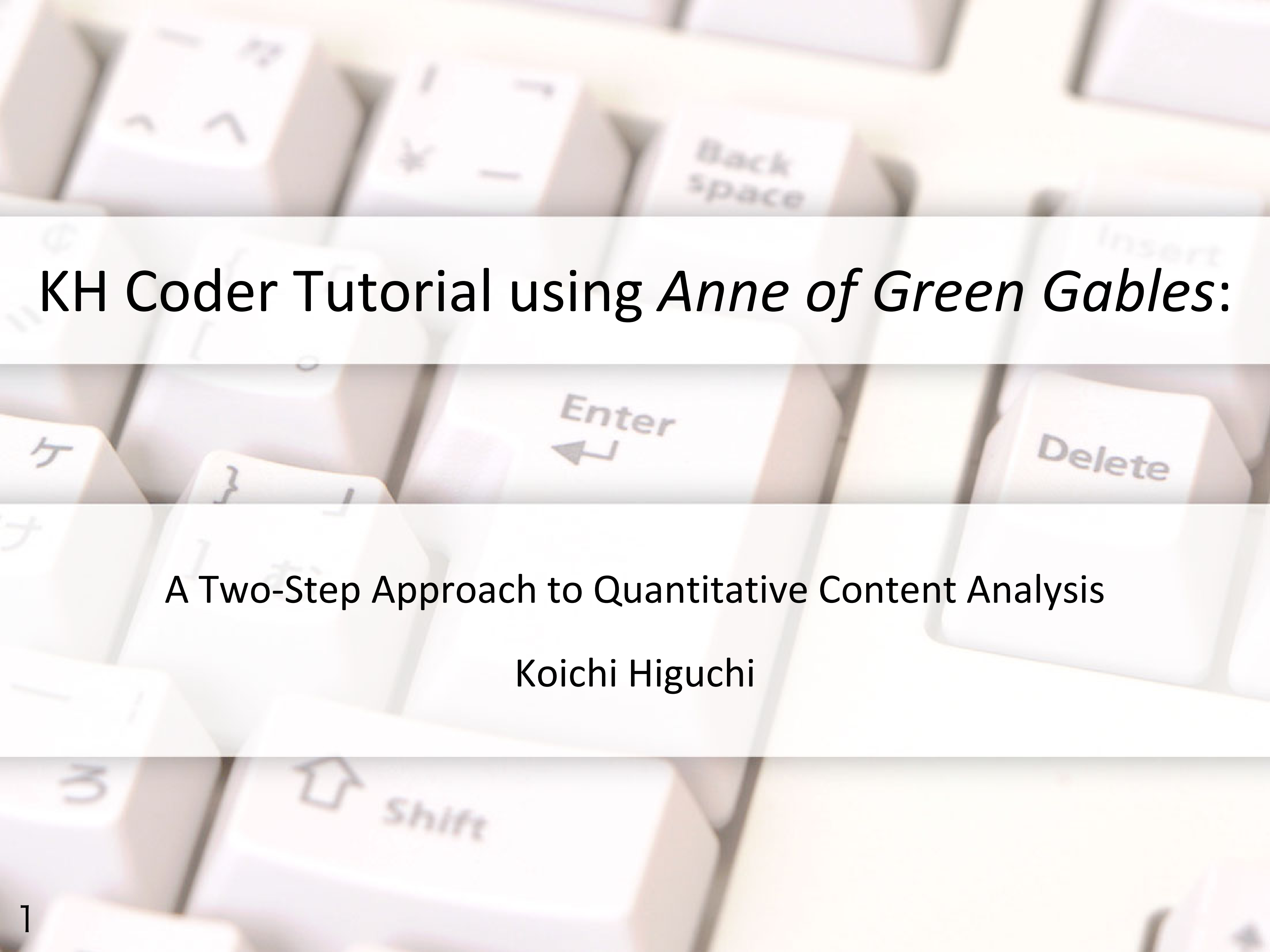


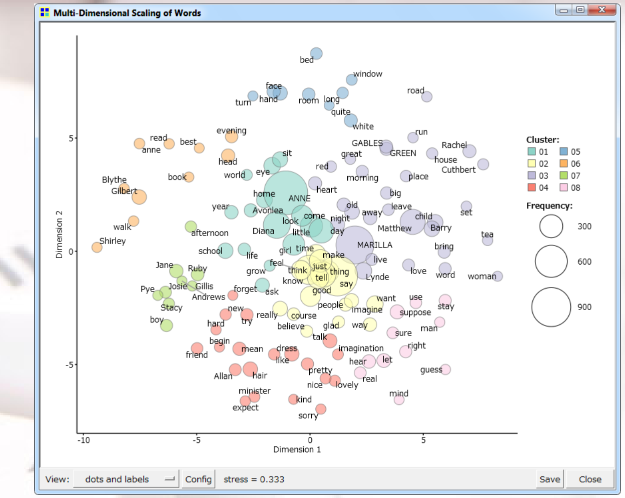
KH Coder Tutorial using *Anne of Green Gables*:

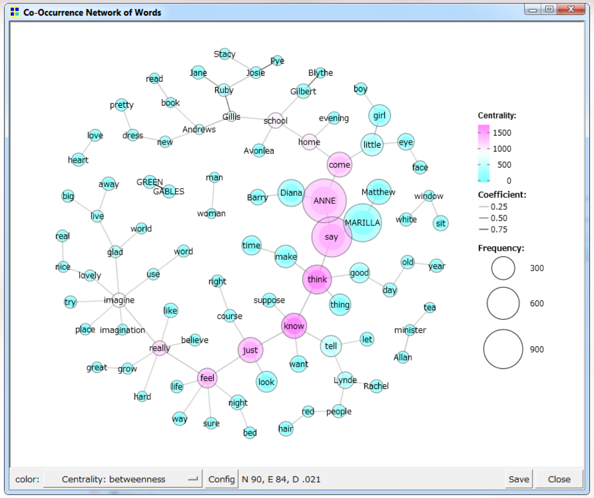
A Two-Step Approach to Quantitative Content Analysis

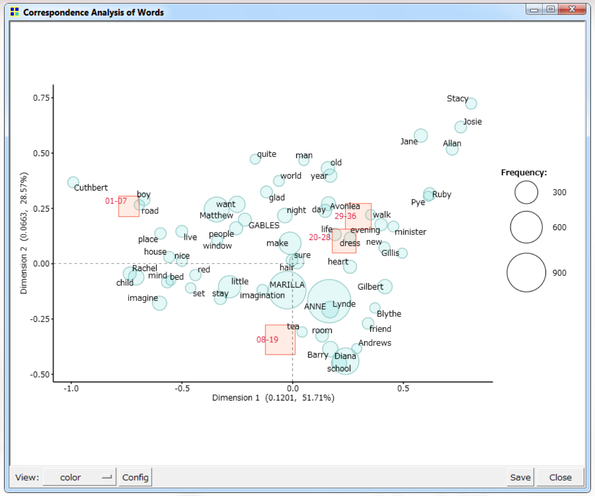
Koichi Higuchi

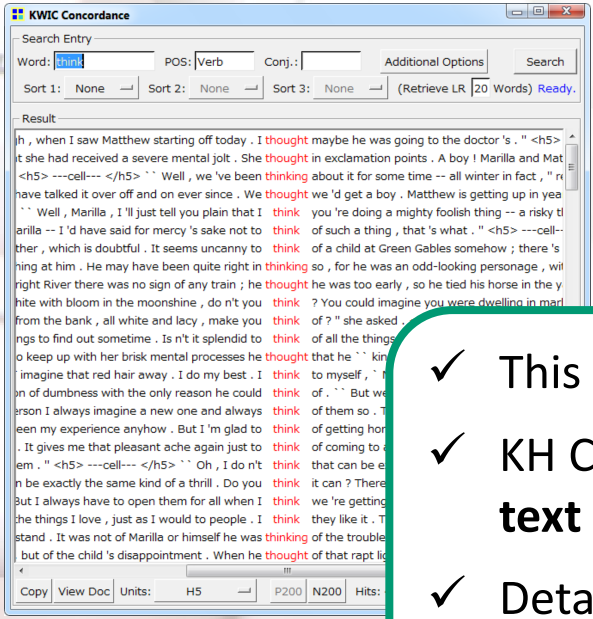
Introduction

Preface






Correspondence Analysis of Words
Dimension 1 (0.1201, 51.71%)
Dimension 2 (0.0663, 28.57%)
Frequency:
View: color
Config
Save
Close

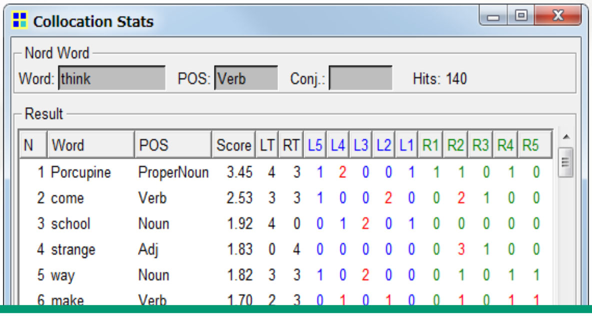


Word: think POS: Verb Conj.:

Sort 1: None Sort 2: None Sort 3: None (Retrieve LR 20 Words) Ready.

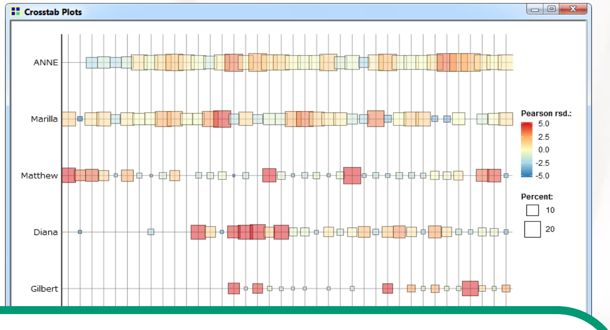
h , when I saw Matthew starting off today . I thought maybe he was going to the doctor 's . " <h5>
t she had received a severe mental jolt . She thought in exclamation points . A boy ! Marilla and Mat
<h5> ---cell--- </h5> `` Well , we 've been thinking about it for some time -- all winter in fact , " re
have talked it over off and on ever since . We thought we 'd get a boy . Matthew is getting up in yea
`` Well , Marilla , I 'll just tell you plain that I think you 're doing a mighty foolish thing -- a risky tl
arilla -- I 'd have said for mercy 's sake not to think of such a thing , that 's what . " <h5> ---cell--
ther , which is doubtful . It seems uncanny to think of a child at Green Gables somehow ; there 's
hing at him . He may have been quite right in thinking so , for he was an odd-looking personage , wi
right River there was no sign of any train ; he thought he was too early , so he tied his horse in the y
hite with bloom in the moonshine , do n't you think ? You could imagine you were dwelling in marl
from the bank , all white and lacy , make you think of ? " she asked .
ngs to find out sometime . Is n't it splendid to think of all the things
o keep up with her brisk mental processes he thought that he `` kin
imagine that red hair away . I do my best . I think to myself , `
on of dumbness with the only reason he could think of . `` But we
rson I always imagine a new one and always think of them so . T
een my experience anyhow . But I 'm glad to think of getting hor
. It gives me that pleasant ache just to think of coming to
em . " <h5> ---cell--- </h5> `` Oh , I do n't think that can be e
n be exactly the same kind of a thrill . Do you think it can ? There
But I always have to open them for all when I think we 're getting
he things I love , just as I would to people . I think they like it . T
stand . It was not of Marilla or himself he was thinking of the trouble
but of the child 's disappointment . When he thought of that rapt li

Units: H5 Hits:



Word: think POS: Verb Conj.: Hits: 140

N	Word	POS	Score	LT	RT	L5	L4	L3	L2	L1	R1	R2	R3	R4	R5
1	Porcupine	ProperNoun	3.45	4	3	1	2	0	0	1	1	1	0	1	0
2	come	Verb	2.53	3	3	1	0	0	2	0	0	2	1	0	0
3	school	Noun	1.92	4	0	0	1	2	0	1	0	0	0	0	0
4	strange	Adj	1.83	0	4	0	0	0	0	0	0	3	1	0	0
5	way	Noun	1.82	3	3	1	0	2	0	0	0	1	0	1	1
6	make	Verb	1.70	2	3	0	1	0	1	0	0	1	0	1	1



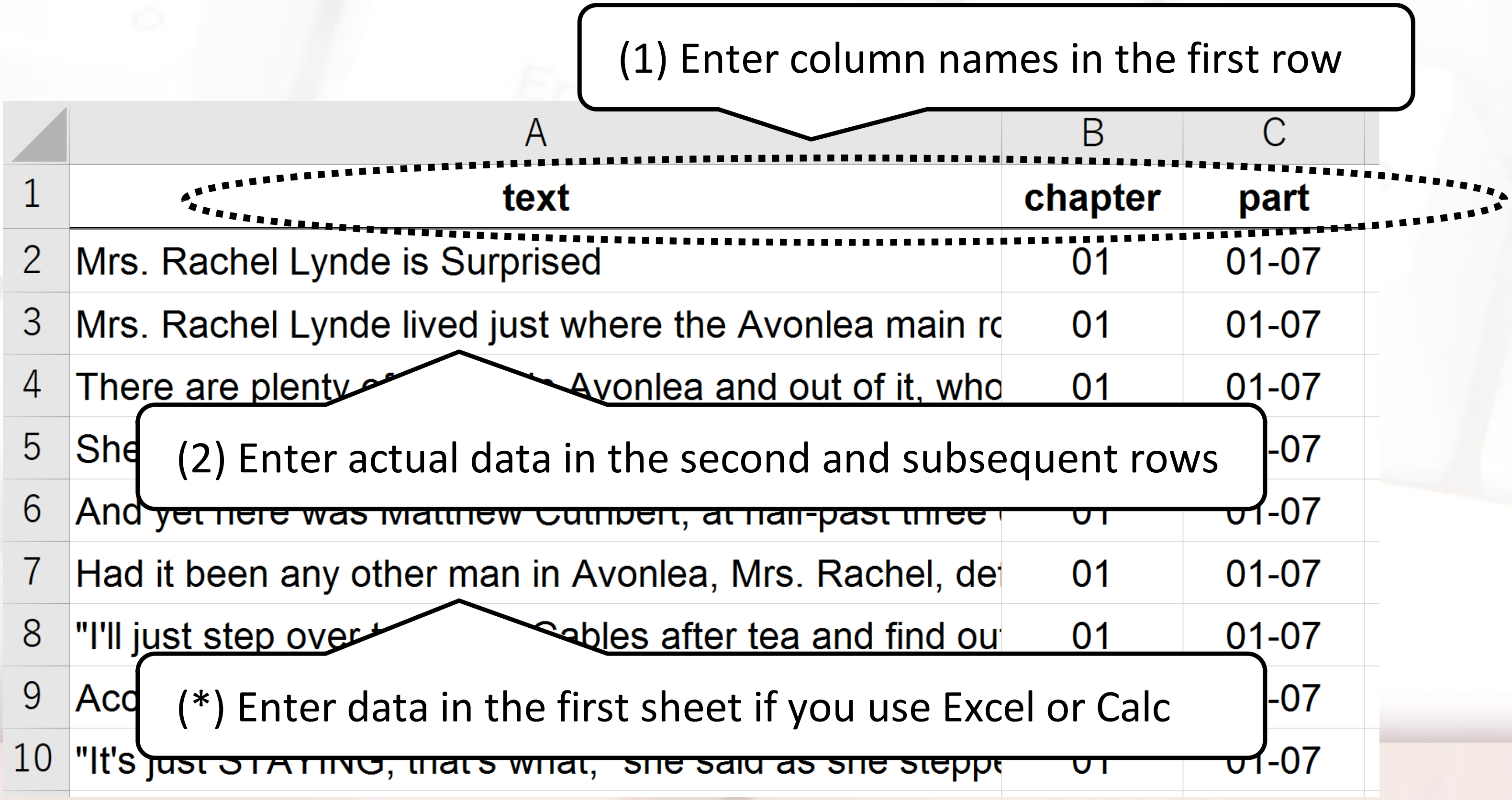
- ✓ This presentation is a tutorial on how to use **KH Coder**.
- ✓ KH Coder is a free software for **quantitative content analysis** or **text mining**. It is also utilized for **computational linguistics**.
- ✓ Details and downloads: http://khcoder.net/en

Table of Contents

- Introduction
  - ✓ Data
  - ✓ Purpose of Analysis
- Preparation
  - ✓ Install KH Coder
  - ✓ Configure Stopwords
  - ✓ Create a Project and Run Pre-Processing
- Step 1
  - ✓ Word Frequency List
  - ✓ The Context where a word is used
  - ✓ Co-occurrence Network of Words
  - ✓ Correspondence Analysis of Words
  - ✓ Closing Remarks for Step 1
- Step 2
  - ✓ Use Coding Rules to Count Concepts
  - ✓ Retrieve Documents Assigned a Specific Code
  - ✓ Characters in Each Chapter
  - ✓ Characters and Verbs
  - ✓ Change of Words Co-occurring with Marilla
  - ✓ Conclusions

Data

- We are going to analyze a novel *Anne of Green Gables* by Montgomery.
- When you prepare your own data for analysis, please open the attached "Anne.xls" file in "tutorial_en" folder and see the figure below.



(1) Enter column names in the first row

(2) Enter actual data in the second and subsequent rows

(*) Enter data in the first sheet if you use Excel or Calc

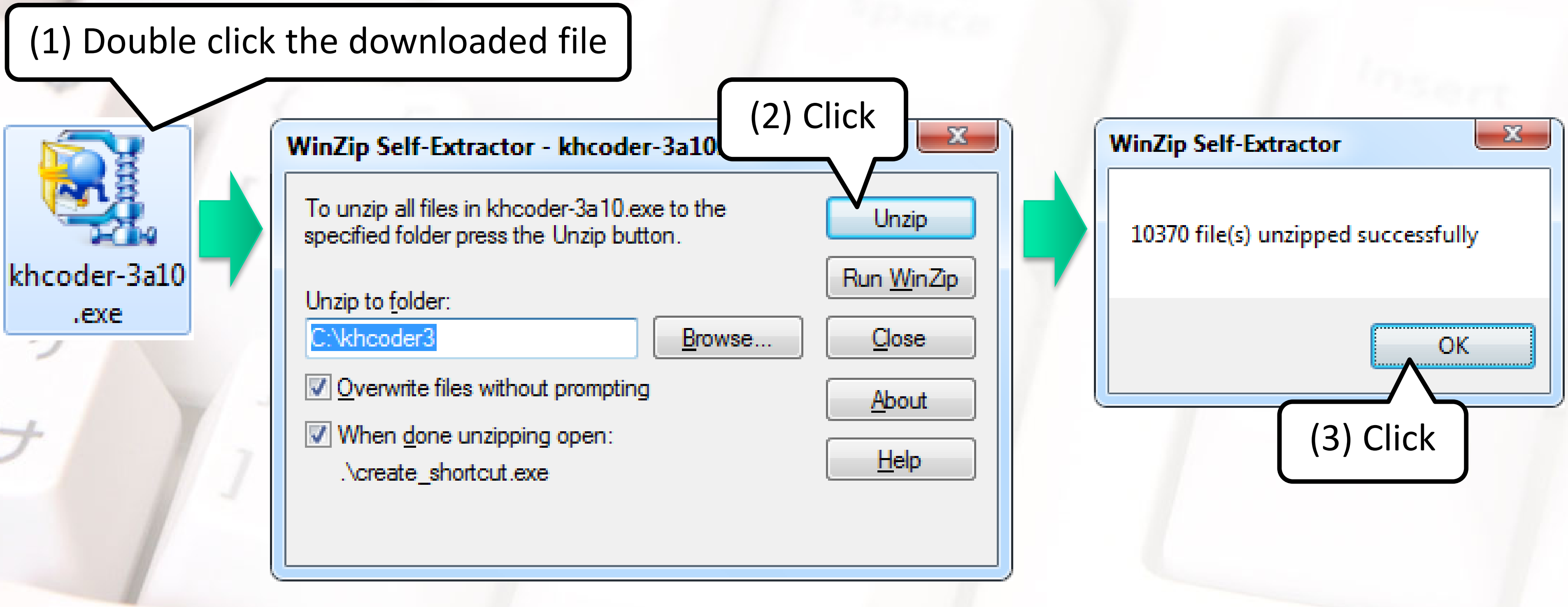
	A	B	C
1	**text**	**chapter**	**part**
2	Mrs. Rachel Lynde is Surprised	01	01-07
3	Mrs. Rachel Lynde lived just where the Avonlea main rc	01	01-07
4	There are plenty o Avonlea and out of it, who	01	01-07
5	She		-07
6	And yet here was Matthew Cuthbert, at half-past three	01	01-07
7	Had it been any other man in Avonlea, Mrs. Rachel, de	01	01-07
8	"I'll just step over t Gables after tea and find ou	01	01-07
9	Acc		-07
10	"It's just STAYING, that's what," she said as she steppe	01	01-07

Purpose of Analysis

- To confirm whether the quantitative analysis can also illustrate the centrality of Marilla
  - ✓ It has been pointed out that the heroine Anne’s foster mother Marilla plays an essential role in the novel and that Marilla is more central than Anne's best friend Diana, and Gilbert with whom Anne has a faint romance.
- To demonstrate a quantitative content analysis approach that comprises the following two steps:
  - ✓ [Step 1] Extract words automatically from data and statistically analyze them to obtain a whole picture and explore the features of the data while avoiding the prejudices of the researcher.
  - ✓ [Step 2] Specify coding rules, such as "if there is a particular expression, we regard it as an appearance of the concept A", and extract concepts from the data. Then, statistically analyze the concepts to deepen the analysis.

Preparation

Install KH Coder



(1) Double click the downloaded file

khcoder-3a10.exe

(2) Click

WinZip Self-Extractor - khcoder-3a10

To unzip all files in khcoder-3a10.exe to the specified folder press the Unzip button.

Unzip to folder: C:\khcoder3

☑ Overwrite files without prompting

☑ When done unzipping open: .\create_shortcut.exe

WinZip Self-Extractor

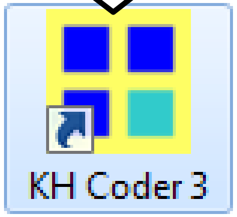
10370 file(s) unzipped successfully

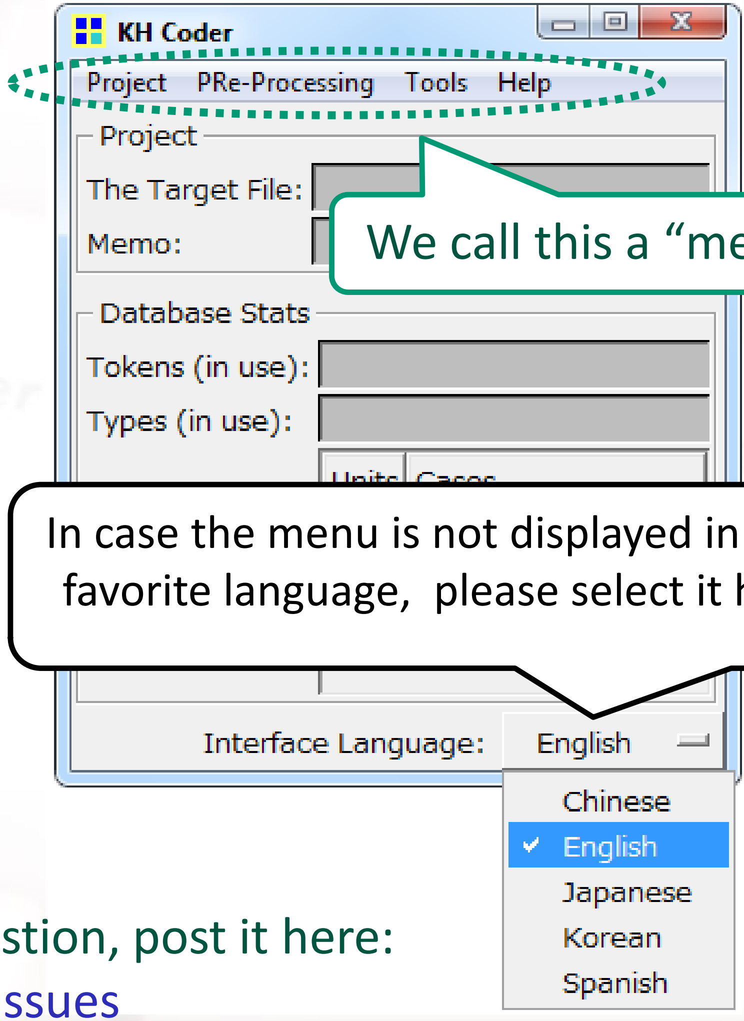
(3) Click

- Now you are ready.
- The number of unzipped files may vary between versions.

Interface Language

(1) Double click the shortcut on your desktop to start KH Coder





Interface Language: English

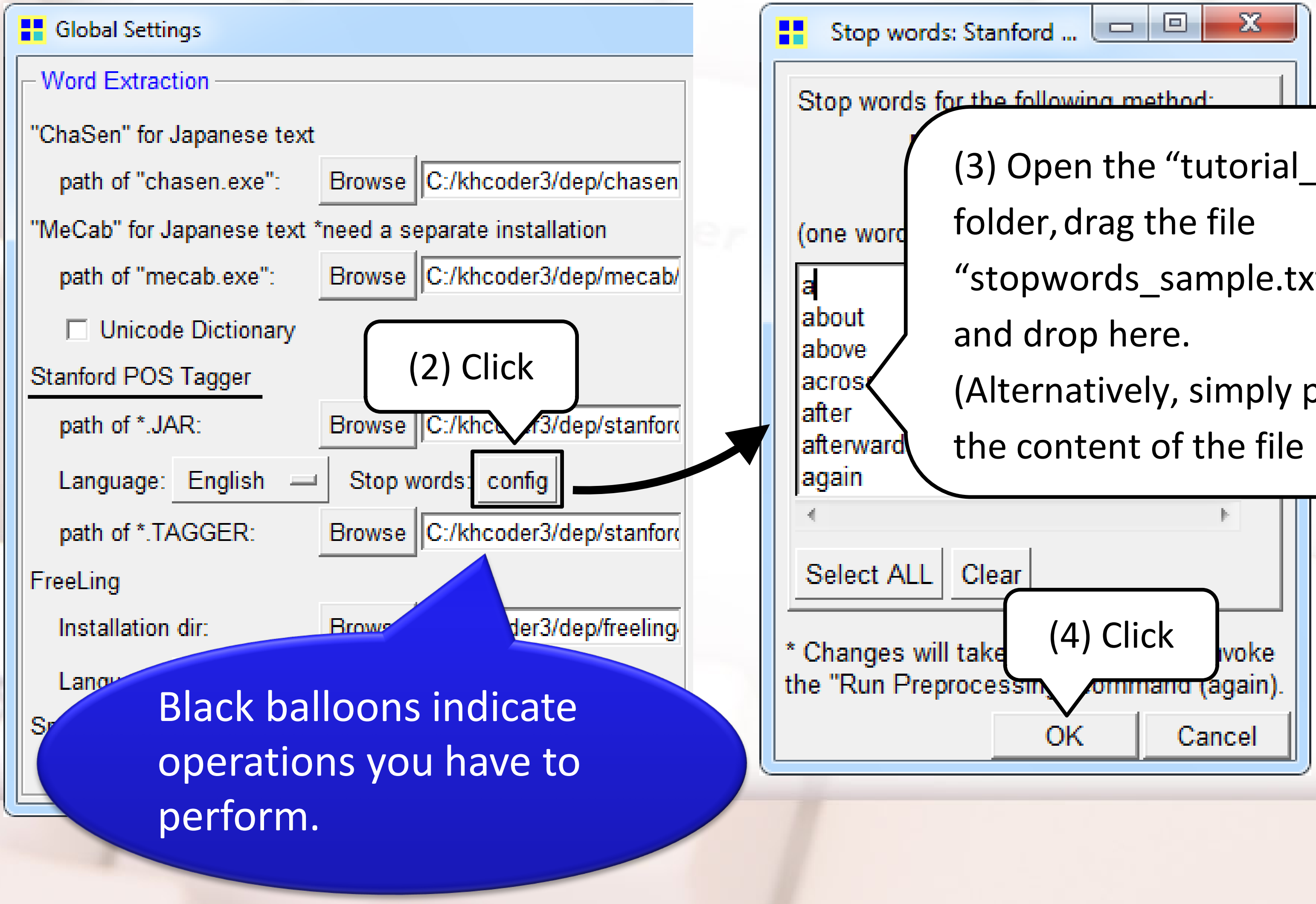
We call this a “menu”.

In case the menu is not displayed in your favorite language, please select it here

- Interface translation is not completed.
- If you find a typo or if you have a suggestion, post it here: https://github.com/ko-ichi-h/khcoder/issues

Configure Stopwords

(1) Go to [Project] [Settings] in the menu of KH Coder



Global Settings

Word Extraction

"ChaSen" for Japanese text

path of "chasen.exe": Browse C:/khcoder3/dep/chasen

"MeCab" for Japanese text *need a separate installation

path of "mecab.exe": Browse C:/khcoder3/dep/mecab/

□ Unicode Dictionary

Stanford POS Tagger

path of *.JAR: Browse C:/khc ... r3/dep/stanford

Language: English Stop words: config

path of *.TAGGER: Browse C:/khcoder3/dep/stanford

FreeLing

Installation dir: Brow ... der3/dep/freeling

(2) Click

Stop words: Stanford ...

Stop words for the following method:

(one word ...

a
about
above
acros
after
afterward
again

Select ALL Clear

* Changes will take ... voke the "Run Preprocessin... ommand (again).

OK Cancel

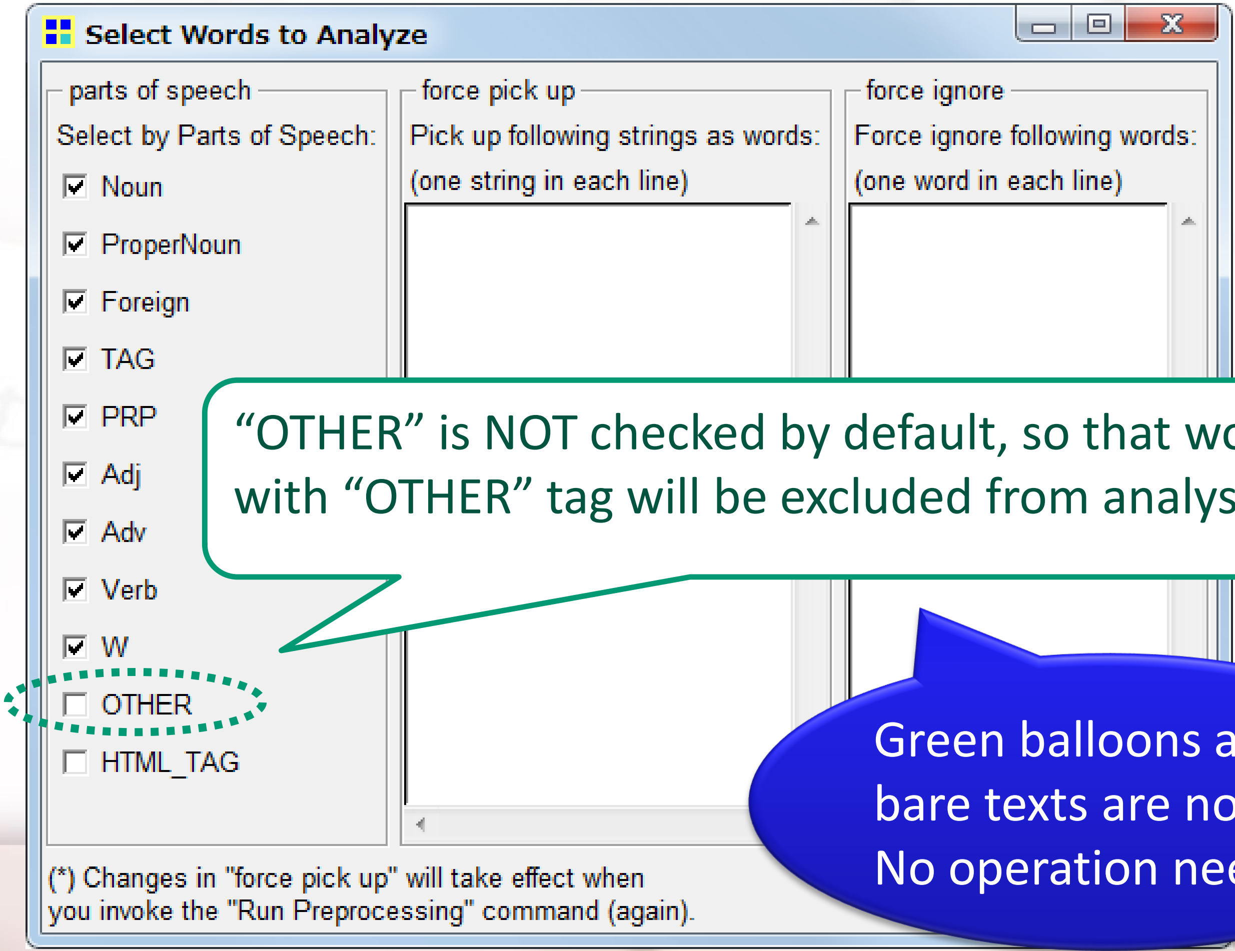
(3) Open the "tutorial_en" folder, drag the file "stopwords_sample.txt" and drop here. (Alternatively, simply paste the content of the file here.)

(4) Click

Black balloons indicate operations you have to perform.

Notes on Stopwords

- You can specify any words as stopwords in KH Coder to exclude those words from your analysis.
- Stopwords will be given the special POS tag “OTHER”.



Select Words to Analyze

parts of speech

Select by Parts of Speech:

☑ Noun
☑ ProperNoun
☑ Foreign
☑ TAG
☑ PRP
☑ Adj
☑ Adv
☑ Verb
☑ W
☐ OTHER
☐ HTML_TAG

force pick up

Pick up following strings as words:
(one string in each line)

force ignore

Force ignore following words:
(one word in each line)

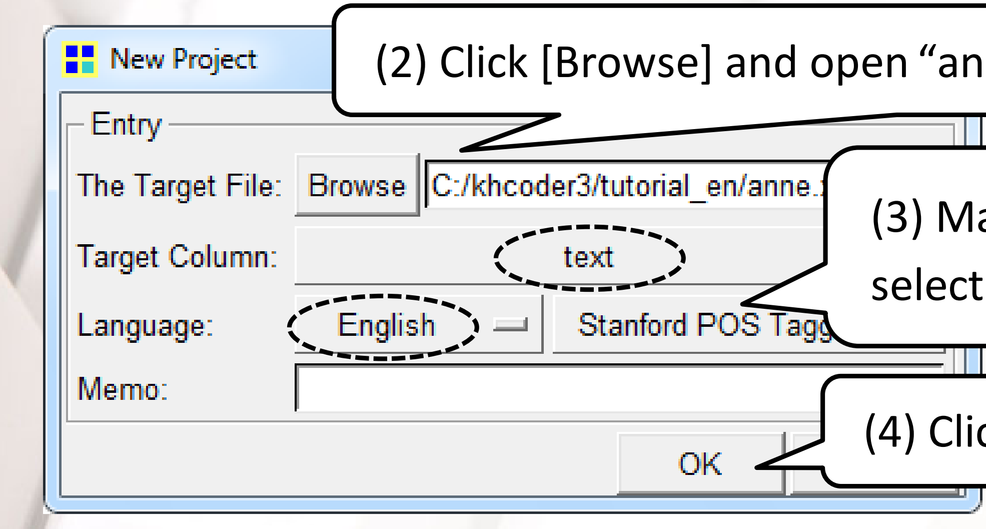
(*) Changes in "force pick up" will take effect when you invoke the "Run Preprocessing" command (again).

“OTHER” is NOT checked by default, so that words with “OTHER” tag will be excluded from analyses.

Green balloons and bare texts are notes. No operation needed.

Create a Project & Run Pre-Processing

(1) Go to [Project] [New] in the menu of KH Coder



The Target File: Browse C:/khcoder3/tutorial_en/anne.x

Target Column: text

Language: English Stanford POS Tagg

Memo:

(2) Click [Browse] and open “anne.xls” in the “tutorial_en” folder

(3) Make sure [text] and [English] are selected

(4) Click

(5) Go to [Pre-Processing] [Run Pre-Processing] in the menu and click [OK]

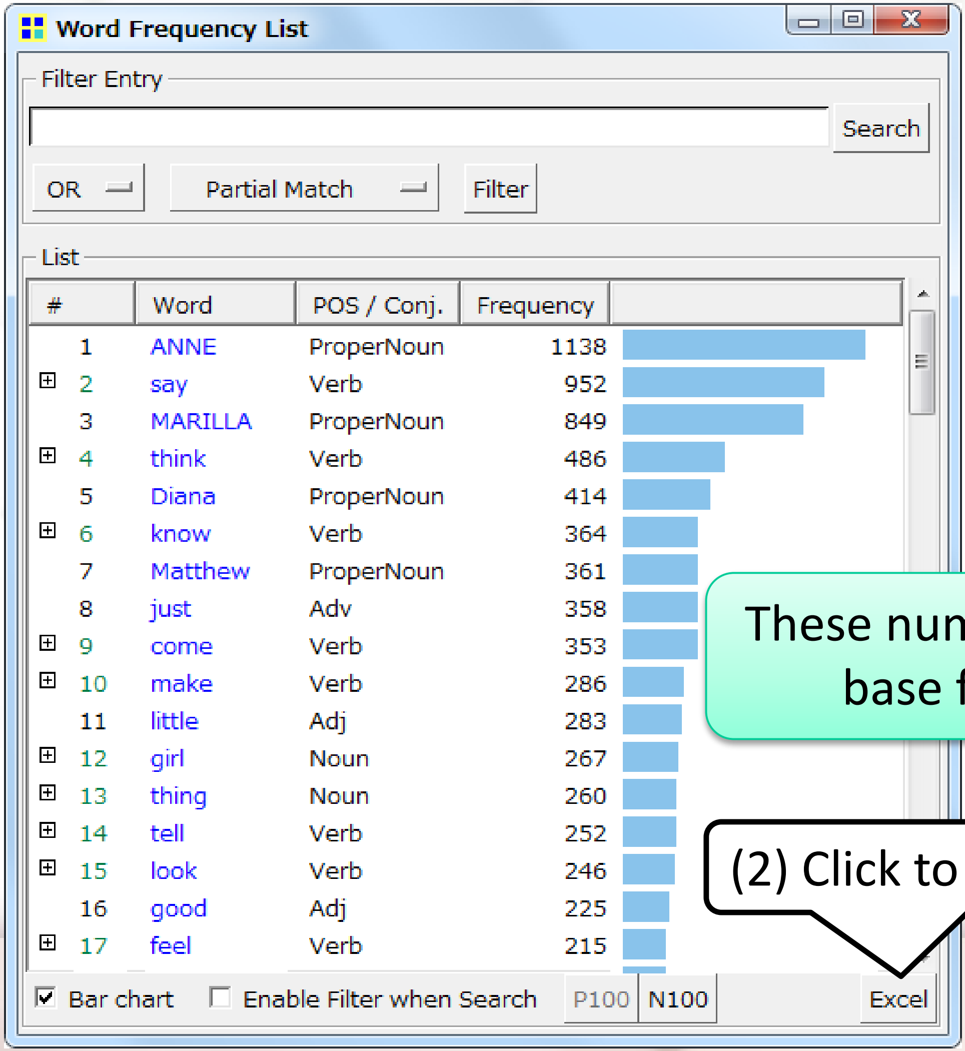
- Next time you start KH Coder, go to [Project] [Open] in the menu and open the project you have created here.
- KH Coder “concentrates” on the task. So it may look frozen or “not responding”. But it’s normal when it’s busy.

Step 1

Step 1

Word Frequency List (1/2)

(1) Go to [Tools] [Words] [Frequency List] in the menu



Word Frequency List

Filter Entry

Search

OR | Partial Match | Filter

List

#	Word	POS / Conj.	Frequency
1	ANNE	ProperNoun	1138
⊞ 2	say	Verb	952
3	MARILLA	ProperNoun	849
⊞ 4	think	Verb	486
5	Diana	ProperNoun	414
⊞ 6	know	Verb	364
7	Matthew	ProperNoun	361
8	just	Adv	358
⊞ 9	come	Verb	353
⊞ 10	make	Verb	286
11	little	Adj	283
⊞ 12	girl	Noun	267
⊞ 13	thing	Noun	260
⊞ 14	tell	Verb	252
⊞ 15	look	Verb	246
16	good	Adj	225
⊞ 17	feel	Verb	215

☑ Bar chart ☐ Enable Filter when Search P100 N100 Excel

These numbers are counts of base forms / lemma

(2) Click to export to Excel

Step 1

Word Frequency List (2/2)

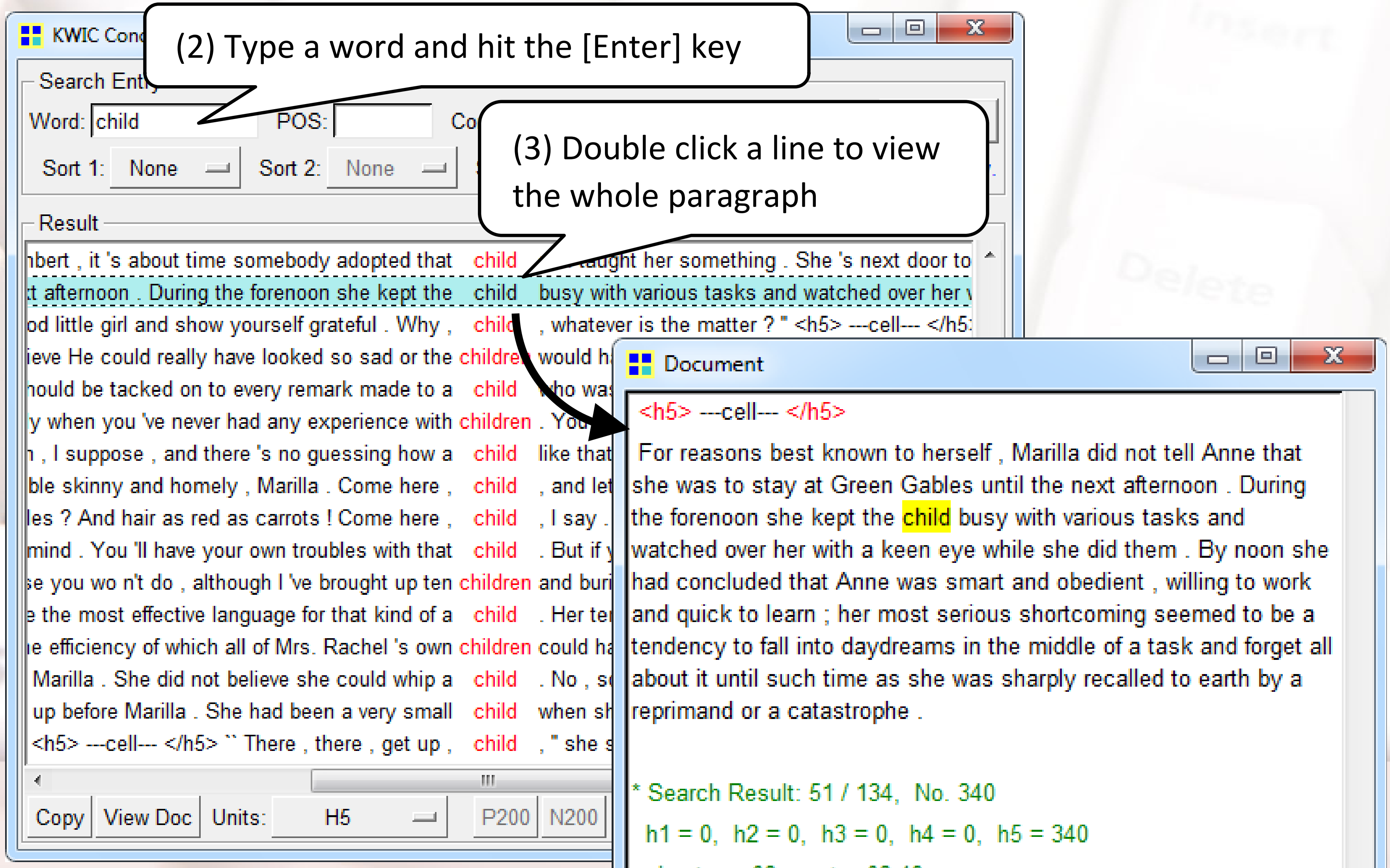
Words	Freq	Words	Freq	Words	Freq
ANNE	1138	little	283	want	149
say	952	girl	267	home	136
MARILLA	849	thing	260	child	134
think	486	tell	252	Barry	132
Diana	414	look	246	school	128
know	364	good	225	sit	126
Matthew	361	feel	215	night	117
just	358	time	208	really	116
come	353	eye	152	hair	114
make	286	Lynde	151	Gilbert	113

- The character name that most frequently appears next to the heroine “ANNE” is not her best friend “Diana” but “MARILLA”.
- In the novel, an orphan “girl” or “child” heroine gets adopted, finds a “home”, and goes to “school”. And she once had a inferiority complex about her “hair”.

Step 1

The Context Where a Word is Used

(1) Go to [Tools] [Words] [KWIC Concordance] in the menu



(2) Type a word and hit the [Enter] key

(3) Double click a line to view the whole paragraph

Search Entry

Word: child POS: Co

Sort 1: None Sort 2: None

Result

```
hbert , it 's about time somebody adopted that   child   taught her something . She 's next door to
t afternoon . During the forenoon she kept the   child   busy with various tasks and watched over her v
od little girl and show yourself grateful . Why ,   child   , whatever is the matter ? " <h5> ---cell--- </h5
ieve He could really have looked so sad or the childre   would h
hould be tacked on to every remark made to a   child   who was
y when you 've never had any experience with children . You
n , I suppose , and there 's no guessing how a   child   like that
ble skinny and homely , Marilla . Come here ,   child   , and let
les ? And hair as red as carrots ! Come here ,   child   , I say .
mind . You 'll have your own troubles with that   child   . But if y
se you wo n't do , although I 've brought up ten children and buri
e the most effective language for that kind of a   child   . Her te
ie efficiency of which all of Mrs. Rachel 's own children could ha
 Marilla . She did not believe she could whip a   child   . No , sc
 up before Marilla . She had been a very small   child   when sh
 <h5> ---cell--- </h5> `` There , there , get up ,   child   , " she s
```

Copy View Doc Units: H5 P200 N200

<h5> ---cell--- </h5>

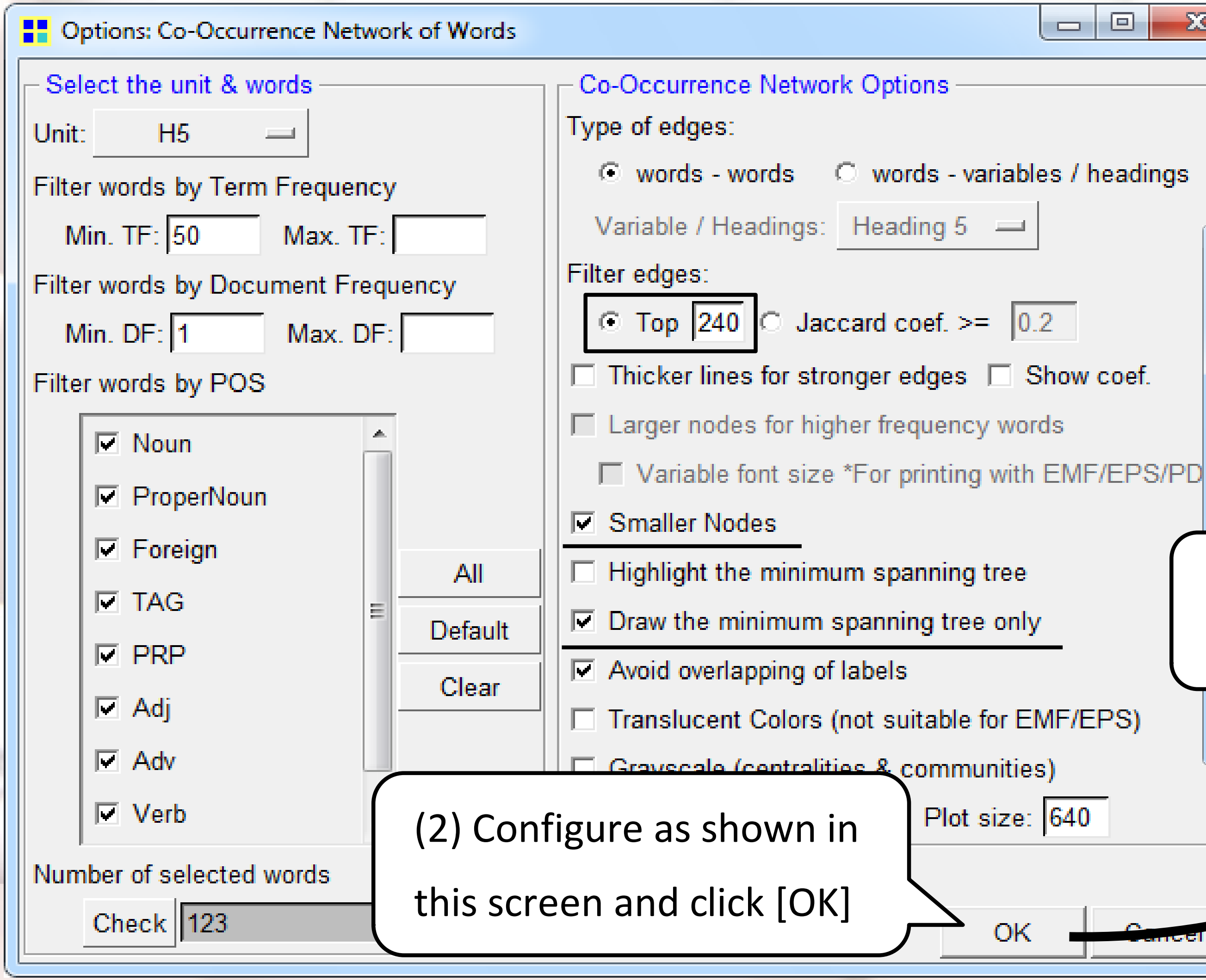
For reasons best known to herself , Marilla did not tell Anne that she was to stay at Green Gables until the next afternoon . During the forenoon she kept the child busy with various tasks and watched over her with a keen eye while she did them . By noon she had concluded that Anne was smart and obedient , willing to work and quick to learn ; her most serious shortcoming seemed to be a tendency to fall into daydreams in the middle of a task and forget all about it until such time as she was sharply recalled to earth by a reprimand or a catastrophe .

* Search Result: 51 / 134, No. 340

h1 = 0, h2 = 0, h3 = 0, h4 = 0, h5 = 340

Co-Occurrence Network of Words (1/2)

(1) Go to [Tools] [Words] [Co-Occurrence Network] in the menu



Options: Co-Occurrence Network of Words

Select the unit & words

Unit: H5

Filter words by Term Frequency

Min. TF: 50 Max. TF:

Filter words by Document Frequency

Min. DF: 1 Max. DF:

Filter words by POS

- ☑ Noun
- ☑ ProperNoun
- ☑ Foreign
- ☑ TAG
- ☑ PRP
- ☑ Adj
- ☑ Adv
- ☑ Verb

Number of selected words

123

Co-Occurrence Network Options

Type of edges:

● words - words ○ words - variables / headings

Variable / Headings: Heading 5

Filter edges:

● Top 240 ○ Jaccard coef. >= 0.2

☐ Thicker lines for stronger edges ☐ Show coef.

☐ Larger nodes for higher frequency words

☐ Variable font size *For printing with EMF/EPS/PD

☑ Smaller Nodes

☐ Highlight the minimum spanning tree

☑ Draw the minimum spanning tree only

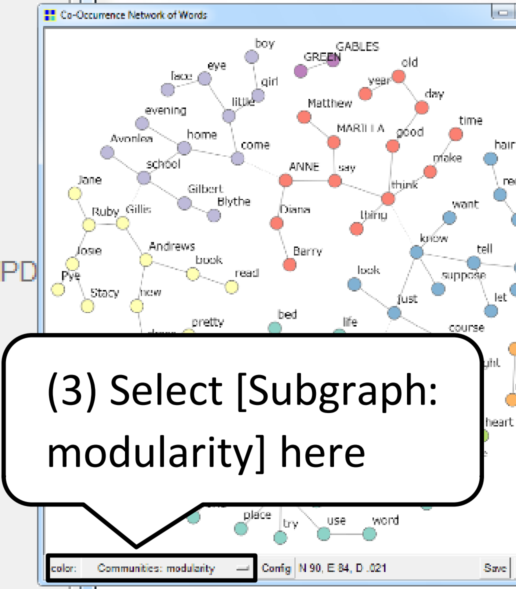
☑ Avoid overlapping of labels

☐ Translucent Colors (not suitable for EMF/EPS)

☐ Grayscale (centralities & communities)

Plot size: 640

(2) Configure as shown in this screen and click [OK]



color: Communities: modularity

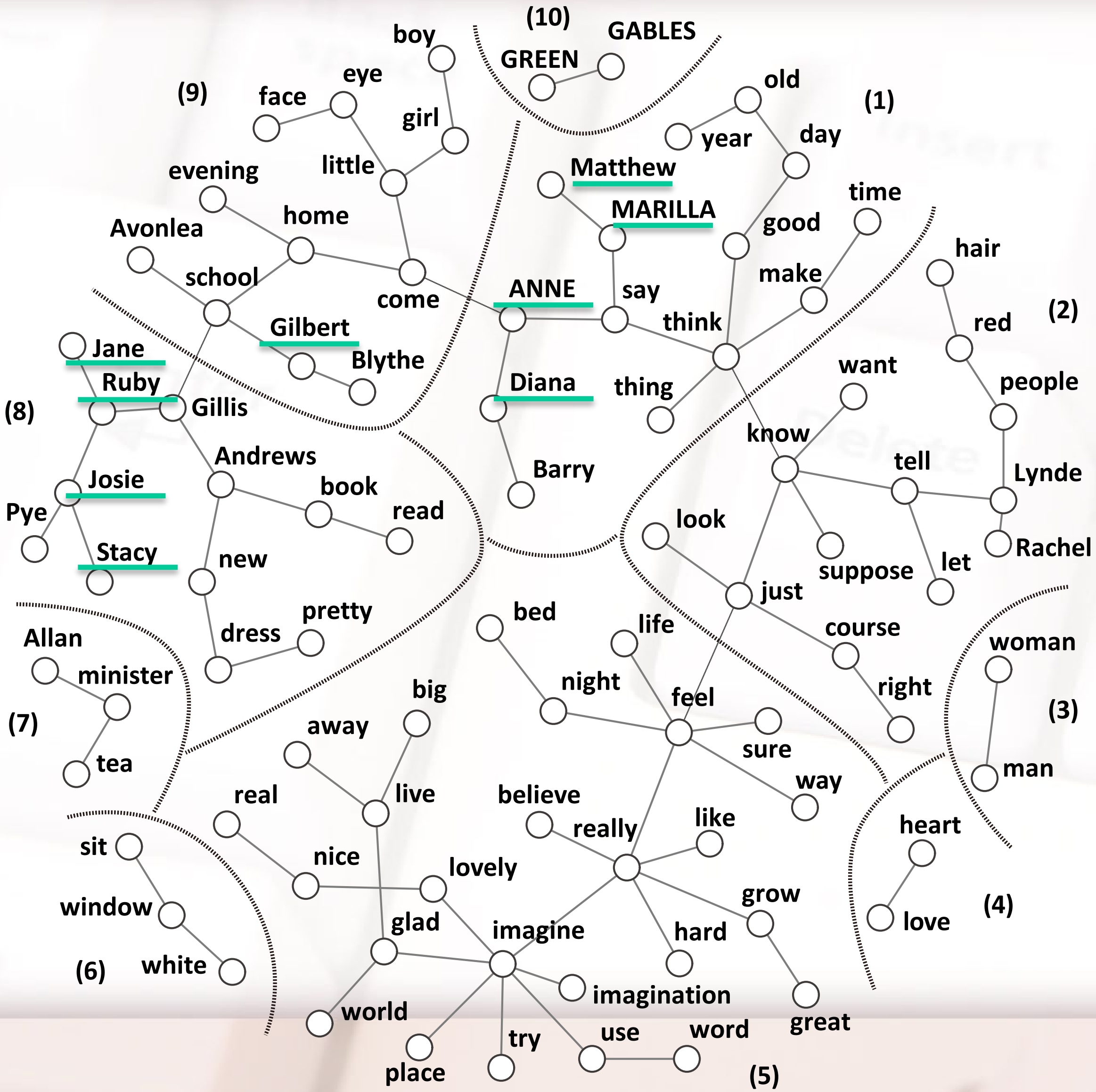
N 90, E 84, D .021

(3) Select [Subgraph: modularity] here

Step 1

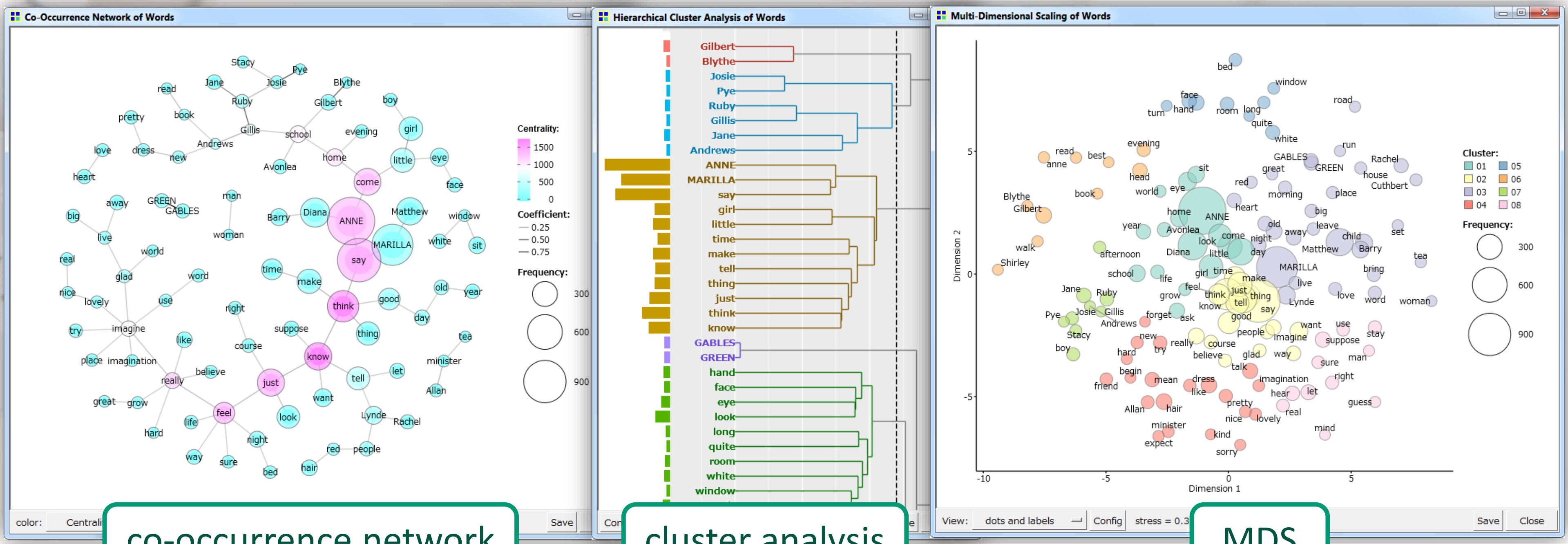
Co-Occurrence Network of Words (2/2)

- “Diana”, “Marilla”, and “Matthew” are connected close to “Anne”
- “Gilbert” is in rather remote part and connected to “Anne” via “school”
- “Jane”, “Ruby”, “Josie”, and “Stacy” are also connected via “school”
- The figure is retouched with Illustrator



Methods for Exploring Co-Occurrences of Words

- To explore co-occurrences of words, you can also use:
  - ✓ hierarchical cluster analysis
  - ✓ Multi-dimensional scaling



co-occurrence network

cluster analysis

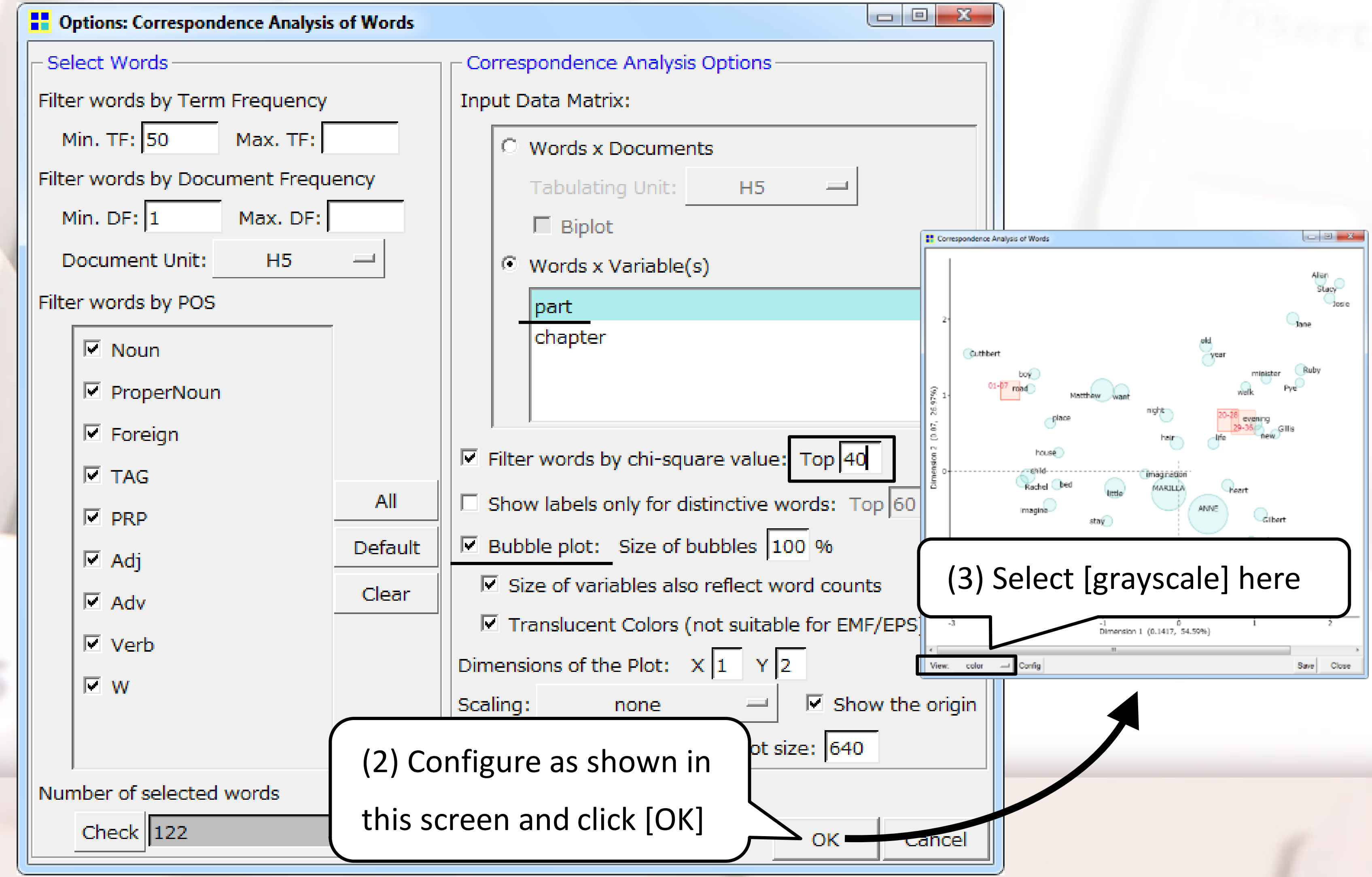
MDS

- By interpreting these result, you may find major themes of the text from groups of words which tend to appear together.
- KH Coder uses R as back end to execute these multivariate methods.

Step 1

Correspondence Analysis of Words (1/2)

(1) Go to [Tools] [Words] [Correspondence Analysis] in the menu



Options: Correspondence Analysis of Words

Select Words

Filter words by Term Frequency

Min. TF: 50 Max. TF:

Filter words by Document Frequency

Min. DF: 1 Max. DF:

Document Unit: H5

Filter words by POS

- ☑ Noun
- ☑ ProperNoun
- ☑ Foreign
- ☑ TAG
- ☑ PRP
- ☑ Adj
- ☑ Adv
- ☑ Verb
- ☑ W

All | Default | Clear

Number of selected words

Check 122

Correspondence Analysis Options

Input Data Matrix:

○ Words x Documents

Tabulating Unit: H5

□ Biplot

● Words x Variable(s)

part
chapter

☑ Filter words by chi-square value: Top 40

□ Show labels only for distinctive words: Top 60

☑ Bubble plot: Size of bubbles 100 %

☑ Size of variables also reflect word counts

☑ Translucent Colors (not suitable for EMF/EPS)

Dimensions of the Plot: X 1 Y 2

Scaling: none ☑ Show the origin

ot size: 640

OK Cancel

(2) Configure as shown in this screen and click [OK]

Correspondence Analysis of Words

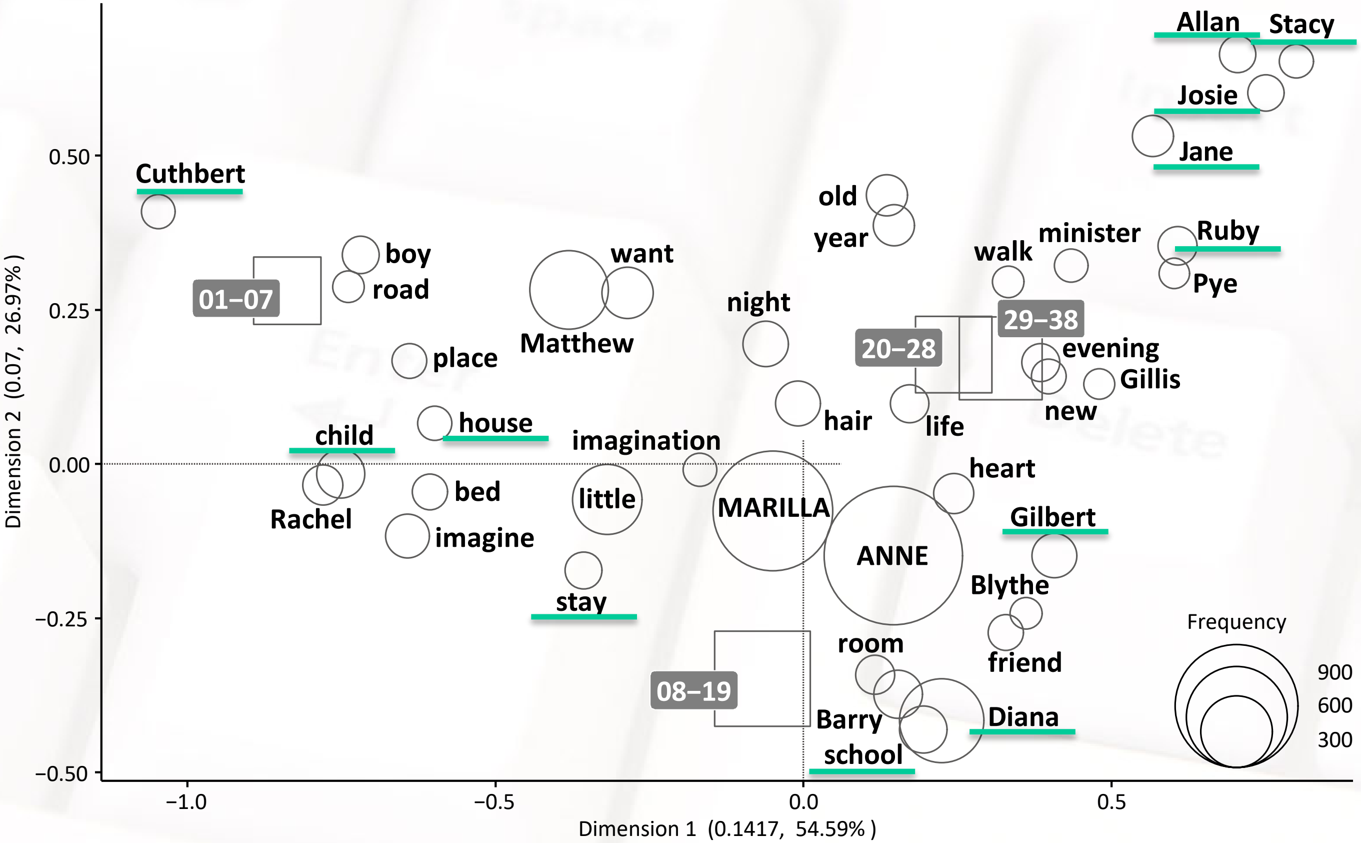
View: color Config Save Close

(3) Select [grayscale] here

Step 1

Correspondence Analysis of Words (2/2)

- In the beginning [01-07], the “child” Anne was allowed to “stay” in “Cuthbert’s house”.
- Then in [08-19], she met a neighbor girl “Diana” and started going to “school”. At the school, she met “Gilbert”.
- In the latter half of the novel, Anne and Diana went separate ways, and Anne's schoolmates, such as “Josie”, “Jane”, and “Ruby”, become characteristic. Anne also learned a lot from adult women such as Mrs. “Allan” and Miss “Stacy”.

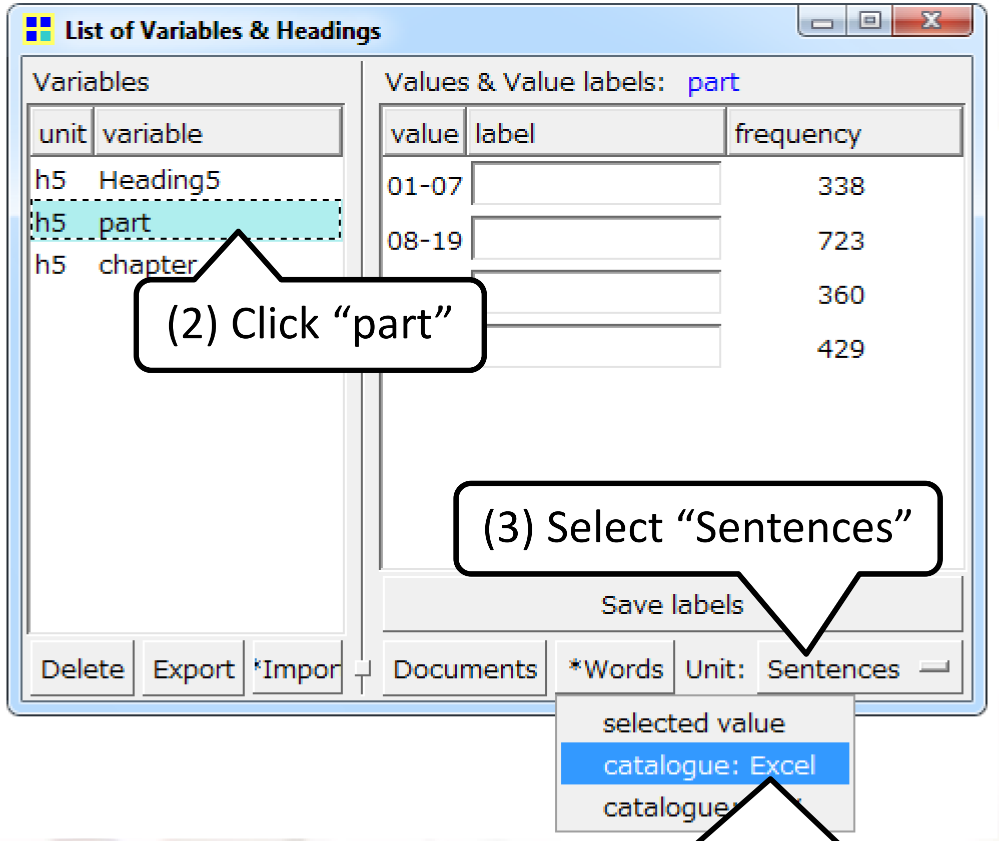


We can understand the story flow throughout the novel by checking characteristic words of each part.

Step 1

Characteristic Words of each Part

(1) Go to [Tools] [Variables & Headings] in the menu



List of Variables & Headings

Variables

unit	variable
h5	Heading5
h5	part
h5	chapter

Values & Value labels: part

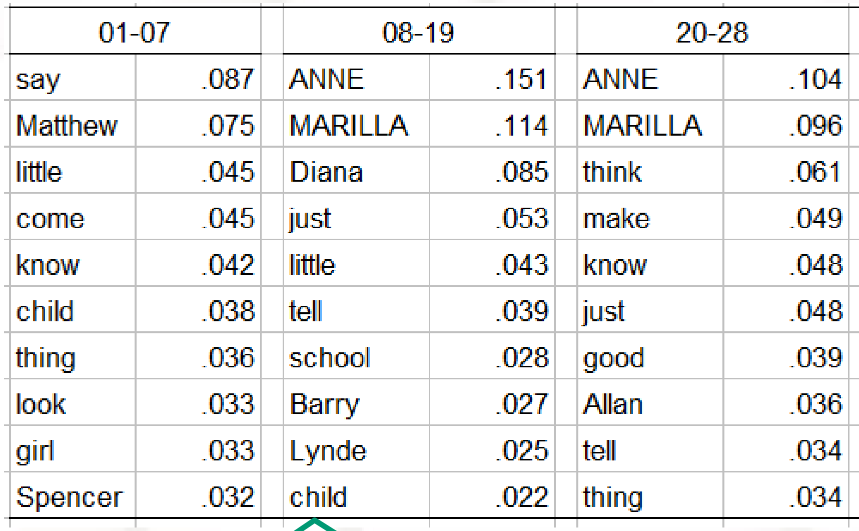
value	label	frequency
01-07		338
08-19		723
		360
		429

Unit: Sentences

(2) Click “part”

(3) Select “Sentences”

(4) Select “catalogue: Excel”



01-07		08-19		20-28	
say	.087	ANNE	.151	ANNE	.104
Matthew	.075	MARILLA	.114	MARILLA	.096
little	.045	Diana	.085	think	.061
come	.045	just	.053	make	.049
know	.042	little	.043	know	.048
child	.038	tell	.039	just	.048
thing	.036	school	.028	good	.039
look	.033	Barry	.027	Allan	.036
girl	.033	Lynde	.025	tell	.034
Spencer	.032	child	.022	thing	.034

Top 10 characteristic words of each part are tabulated. It can be used as an alternative for correspondence analysis.

Step 1

Closing Remarks for Step 1

- Statistical analyses of automatically extracted words are suitable for gaining a whole picture of the data
  - ✓ Main theme (word frequency list or co-occurrence network)
  - ✓ Relations between characters or words (co-occurrence network)
  - ✓ Story flow (correspondence analysis)
- About the centrality of Marilla
  - ✓ Most frequently appears next to the heroine Anne
  - ✓ Her relationship with Anne appears to be almost as strong as Diana’s
  - ✓ Be present throughout all four parts of the story

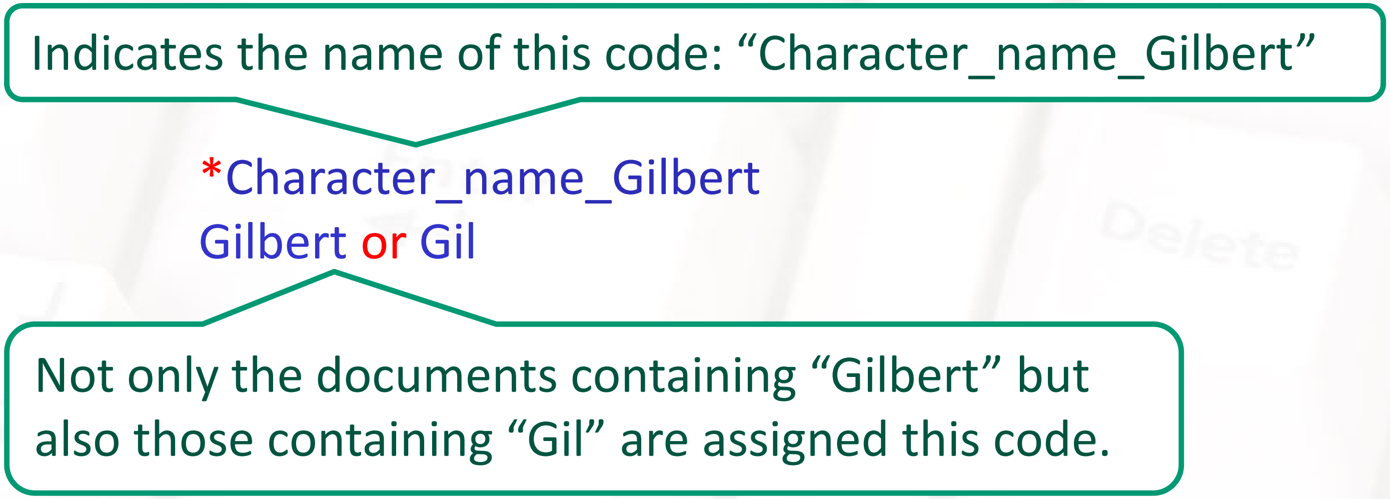
We obtained overviews of entire data in this step. Next, we are going to put more focus on Marilla using coding rules.

Step 2

Step 2

Use Coding Rules to Count Concepts

- In some cases, we have to count concepts, not words.
- To count concepts, you can compose “cording rules” like this:



Indicates the name of this code: “Character_name_Gilbert”

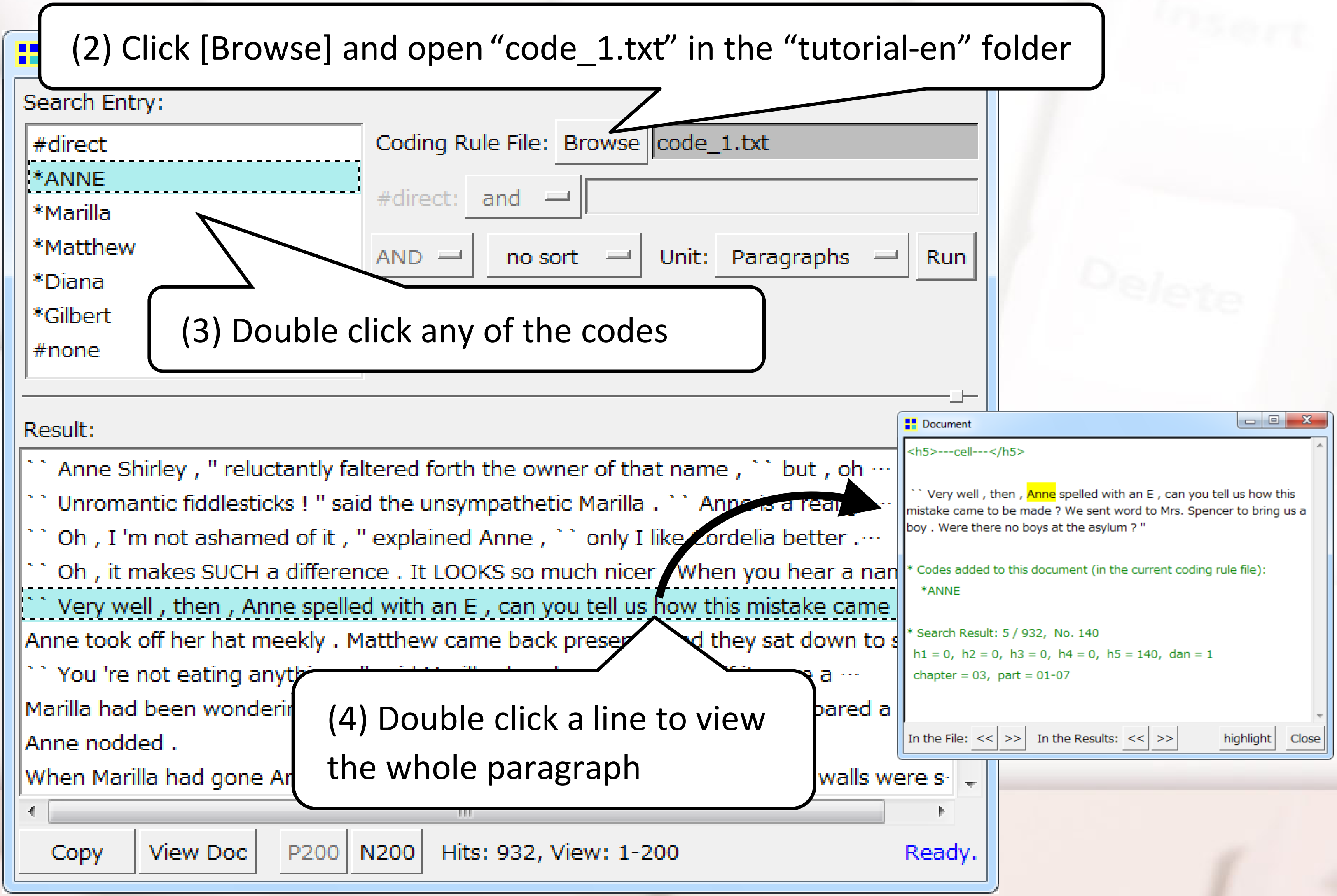
```
*Character_name_Gilbert
Gilbert or Gil
```

Not only the documents containing “Gilbert” but also those containing “Gil” are assigned this code.

- If a document is acceptable under multiple coding rules, multiple codes will be assigned to the document.

Step 2

Retrieve Documents Assigned a Specific Code



(1) Go to [Tools] [Documents] [Search Documents] in the menu

(2) Click [Browse] and open "code_1.txt" in the "tutorial-en" folder

(3) Double click any of the codes

(4) Double click a line to view the whole paragraph

Search Entry:

- #direct
- *ANNE
- *Marilla
- *Matthew
- *Diana
- *Gilbert
- #none

Coding Rule File: code_1.txt

Unit: Paragraphs

Result:

```
`` Anne Shirley , " reluctantly faltered forth the owner of that name , `` but , oh ···
`` Unromantic fiddlesticks ! " said the unsympathetic Marilla . `` Anne is a real
`` Oh , I 'm not ashamed of it , " explained Anne , `` only I like Cordelia better .···
`` Oh , it makes SUCH a difference . It LOOKS so much nicer . When you hear a nan
`` Very well , then , Anne spelled with an E , can you tell us how this mistake came
Anne took off her hat meekly . Matthew came back preser                d they sat down to s
`` You 're not eating anyth                                                    a ···
Marilla had been wonderir                                                 pared a
Anne nodded .
When Marilla had gone Ar                                                  walls were s·
```

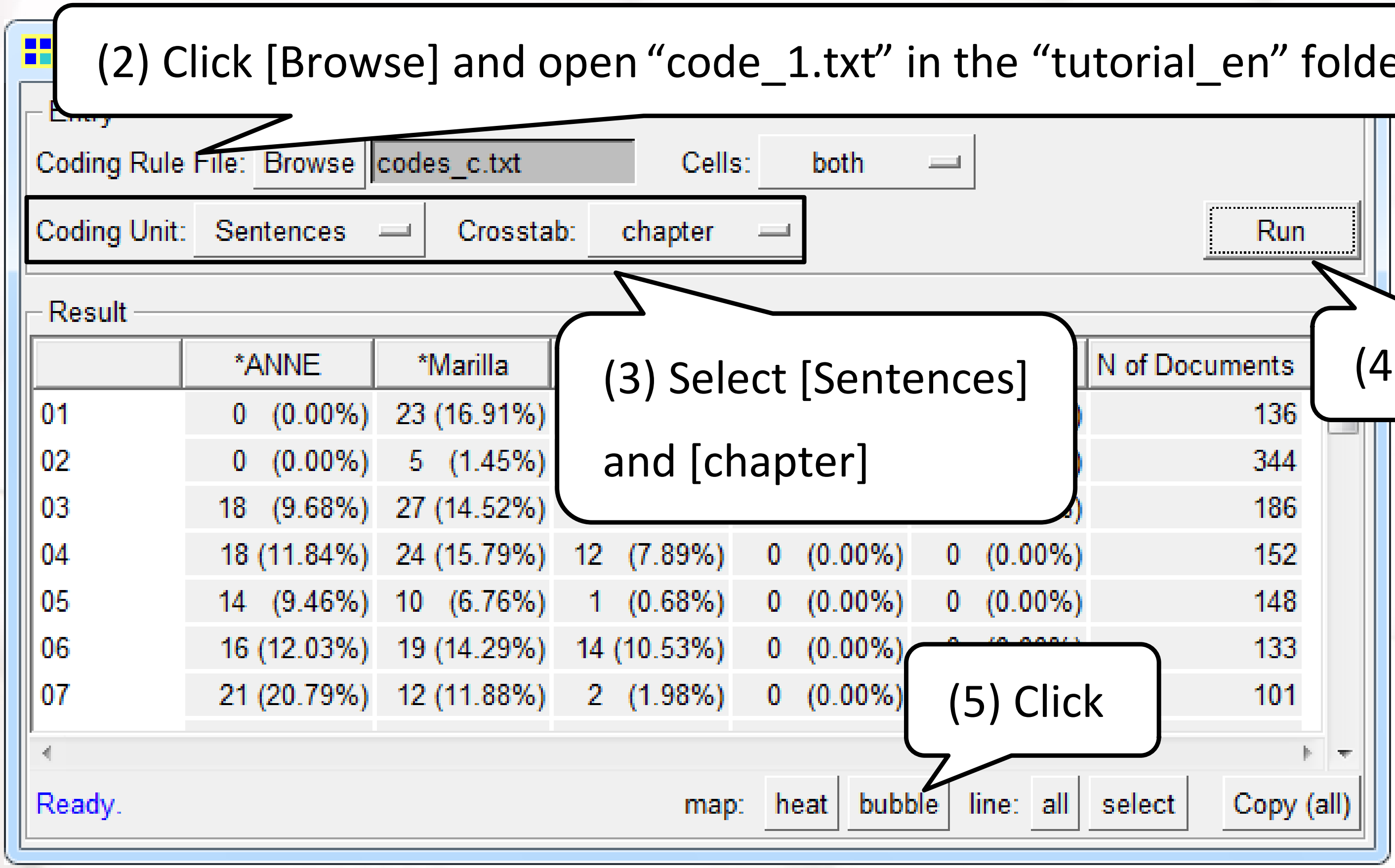
Hits: 932, View: 1-200 Ready.

Document

```
<h5>---cell---</h5>

`` Very well , then , Anne spelled with an E , can you tell us how this mistake came to be made ? We sent word to Mrs. Spencer to bring us a boy . Were there no boys at the asylum ? "

* Codes added to this document (in the current coding rule file):
  *ANNE

* Search Result: 5 / 932,  No. 140
 h1 = 0,  h2 = 0,  h3 = 0,  h4 = 0,  h5 = 140,  dan = 1
 chapter = 03,  part = 01-07
```

Characters in Each Chapter (1/2)

(1) Go to [Tools] [Coding] [Crosstab] in the menu



(2) Click [Browse] and open “code_1.txt” in the “tutorial_en” folder

Coding Rule File: Browse codes_c.txt

Cells: both

Coding Unit: Sentences

Crosstab: chapter

(3) Select [Sentences] and [chapter]

(4) Click

Result

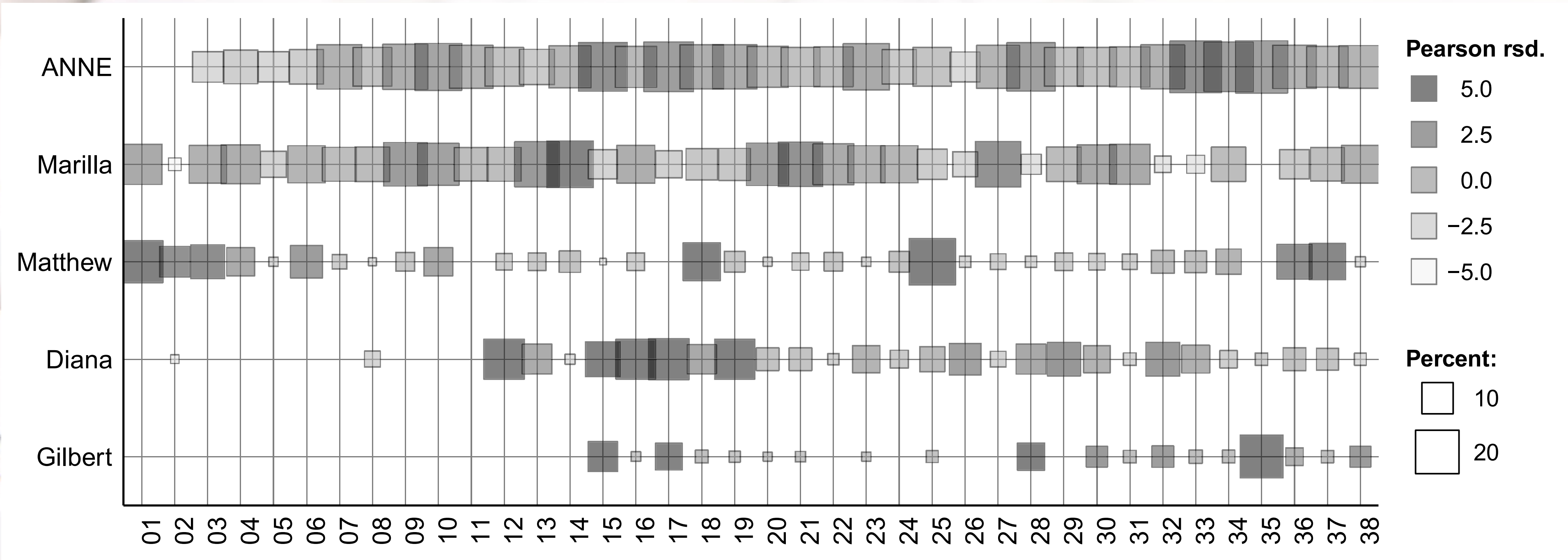
	*ANNE	*Marilla				N of Documents
01	0 (0.00%)	23 (16.91%)				136
02	0 (0.00%)	5 (1.45%)				344
03	18 (9.68%)	27 (14.52%)				186
04	18 (11.84%)	24 (15.79%)	12 (7.89%)	0 (0.00%)	0 (0.00%)	152
05	14 (9.46%)	10 (6.76%)	1 (0.68%)	0 (0.00%)	0 (0.00%)	148
06	16 (12.03%)	19 (14.29%)	14 (10.53%)	0 (0.00%)	0 (0.00%)	133
07	21 (20.79%)	12 (11.88%)	2 (1.98%)	0 (0.00%)		101

Ready.

(5) Click

Characters in Each Chapter (2/2)

Step 2

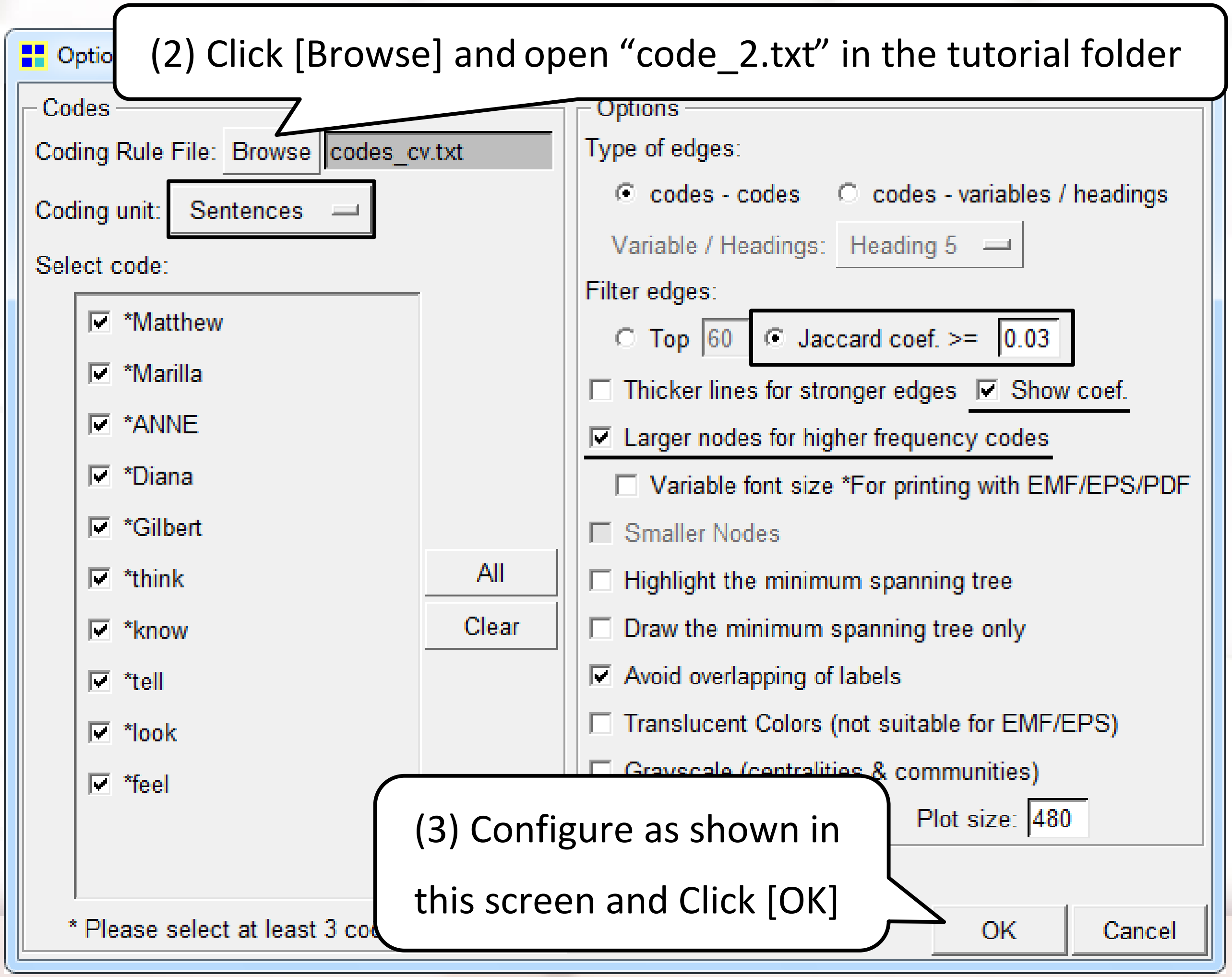


- Marilla and Anne are present almost everywhere
- Although Marilla and Anne were apart in chapter 35, there was an emotional reunion in the following chapter 36. Anne won a scholarship and rejoiced saying “Oh, won’t Matthew and Marilla be pleased!”

Step 2

Characters and Verbs (1/2)

(1) Go to [Tools] [Coding] [Co-occurrences Network] in the menu



(2) Click [Browse] and open “code_2.txt” in the tutorial folder

Optio

Codes

Coding Rule File: Browse codes_cv.txt

Coding unit: Sentences

Select code:

- ☑ *Matthew
- ☑ *Marilla
- ☑ *ANNE
- ☑ *Diana
- ☑ *Gilbert
- ☑ *think
- ☑ *know
- ☑ *tell
- ☑ *look
- ☑ *feel

* Please select at least 3 co

Options

Type of edges:

● codes - codes ○ codes - variables / headings

Variable / Headings: Heading 5

Filter edges:

○ Top 60 ● Jaccard coef. >= 0.03

□ Thicker lines for stronger edges ☑ Show coef.

☑ Larger nodes for higher frequency codes

□ Variable font size *For printing with EMF/EPS/PDF

□ Smaller Nodes

□ Highlight the minimum spanning tree

□ Draw the minimum spanning tree only

☑ Avoid overlapping of labels

□ Translucent Colors (not suitable for EMF/EPS)

□ Grayscale (centralities & communities)

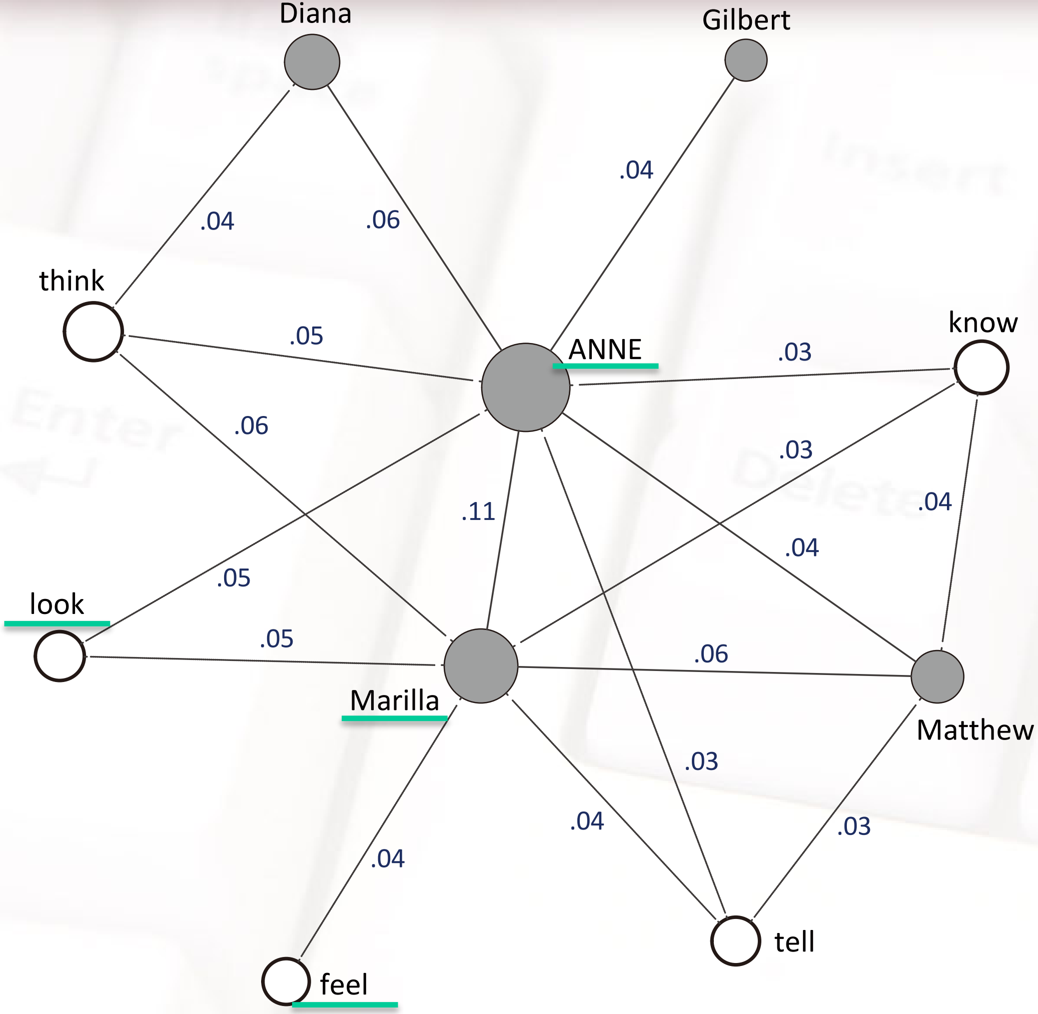
Plot size: 480

(3) Configure as shown in this screen and Click [OK]

Step 2

Characters and Verbs (2/2)

- Anne often expresses what she "feels" to Marilla:
  - ✓ "I do feel dreadfully sad, Marilla" (c21)
- Marilla and Anne often "look" at each other:
  - ✓ Marilla looked at Anne and softened at sight of the child's pale face... (c6)
  - ✓ Anne looked at her with eyes limpid with sympathy (c20)
  - ✓ Marilla looked at her with a tenderness that would never have been suffered to reveal itself in any clearer light... (c30)

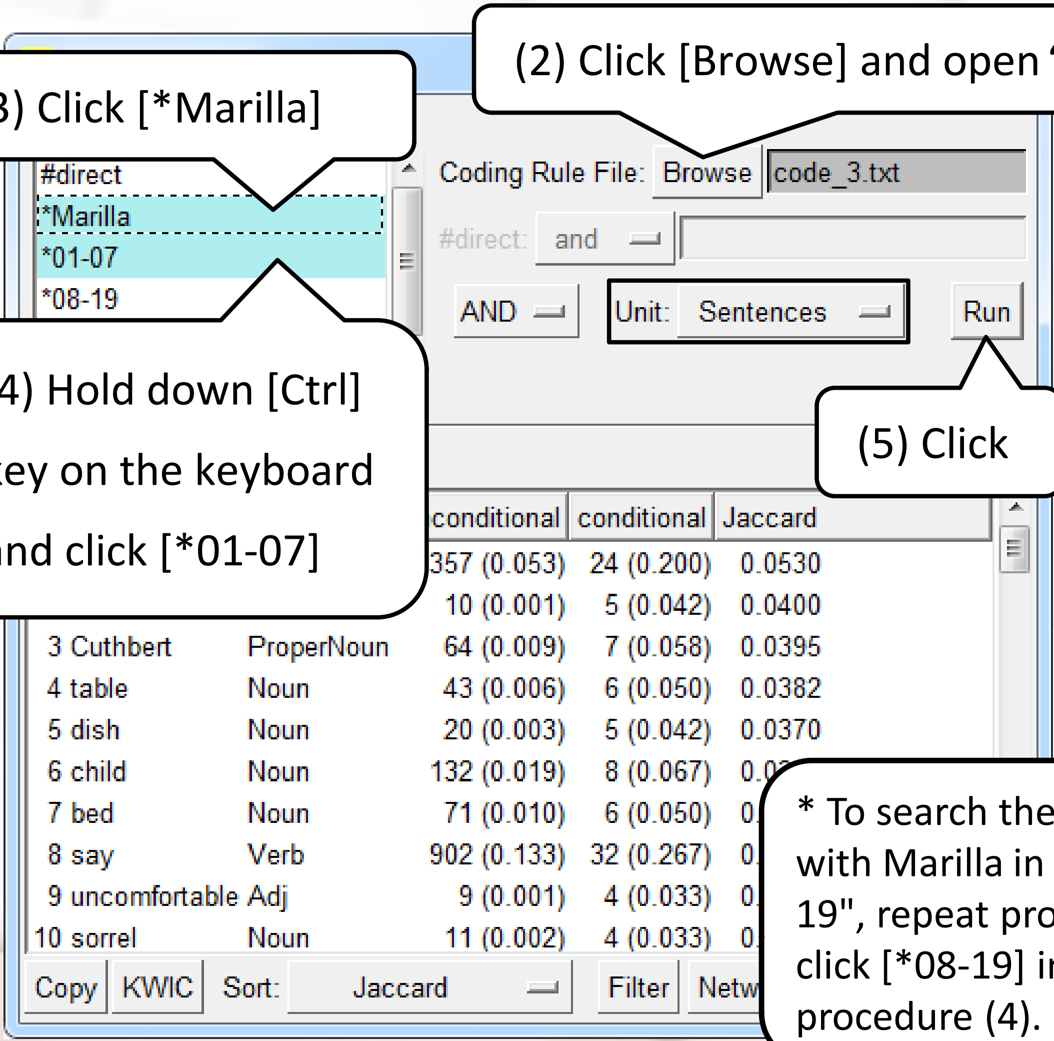


Marilla and Anne exchange their feelings by words, and also with their eyes, meaning that a close and intimate relationship is depicted between the two.

Step 2

Change of Words Co-occurring with Marilla (1/3)

(1) Go to [Tools] [Words] [Word Association] in the menu



#direct
*Marilla
*01-07
*08-19

Coding Rule File: Browse code_3.txt

#direct: and

AND

Unit: Sentences

Run

		conditional	conditional	Jaccard
		357 (0.053)	24 (0.200)	0.0530
		10 (0.001)	5 (0.042)	0.0400
3 Cuthbert	ProperNoun	64 (0.009)	7 (0.058)	0.0395
4 table	Noun	43 (0.006)	6 (0.050)	0.0382
5 dish	Noun	20 (0.003)	5 (0.042)	0.0370
6 child	Noun	132 (0.019)	8 (0.067)	0.0
7 bed	Noun	71 (0.010)	6 (0.050)	0.
8 say	Verb	902 (0.133)	32 (0.267)	0.
9 uncomfortable	Adj	9 (0.001)	4 (0.033)	0.
10 sorrel	Noun	11 (0.002)	4 (0.033)	0.

Copy KWIC Sort: Jaccard Filter Netw

(2) Click [Browse] and open “code_3.txt”

(3) Click [*Marilla]

(4) Hold down [Ctrl] key on the keyboard and click [*01-07]

(5) Click

* To search the words co-occurring with Marilla in the following part "08-19", repeat procedure (3) and then click [*08-19] instead of [*01-07] in procedure (4).

Step 2

Change of Words Co-occurring with Marilla (2/3)

“Marilla really did not know how to talk to the child, and her uncomfortable ignorance made her crisp and...” (c4)

The “feel” and “look”

01-07		08-19		20-28		29-38	
Matthew	.053	say	.072	say	.042	Matthew	.041
mare	.040	ANNE	.059	think	.034	look	.040
Cuthbert	.040	just	.039	ANNE	.032	sit	.039
table	.038	think	.036	cake	.030	ANNE	.038
dish	.037	brooch	.031	make	.028	say	.038
child	.033	tell	.030	minister	.028	face	.031
bed	.032	evening	.025	Allan	.026	girl	.026
say	.032	home	.024	feel	.025	think	.024
uncomfortable	.032	set	.024	know	.024	want	.022
sorrel	.032	let	.023	time	.023	lean	.022

The “child” is upgraded to “Anne” and implying that it is impossible to bring up a child without “saying” anything.

Step 2

Change of Words Co-occurring with Marilla (3/3)

- Change of Marilla

1. Uncomfortable ignorance [01-07]

↓

2. Calling Anne and Saying many things [08-28]

↓

3. Exchanging feelings by words and eyes with Anne [20-38]

The change is depicted throughout the story.

Step 2

Conclusions

- Results of step 2 showed that:
  - ✓ Marilla is literally present almost everywhere
  - ✓ A close and intimate relationship is depicted between Marilla and Anne
  - ✓ Change of Marilla and growing relationship between Marilla and Anne is depicted throughout the story

Our analysis supports the assertion that Marilla plays central roll in the story.

Identifying keywords like “child”, “uncomfortable”, “look”, and “feel” through quantitative analysis is considered to be useful for extracting depiction which specifically describes Marilla’s roll and change in the story.

- Web site of KH Coder

  http://khcoder.net/en

- For more details on this tutorial

  Part 1: http://www.ritsumei.ac.jp/file.jsp?id=325881
  Part 2: http://www.ritsumei.ac.jp/file.jsp?id=346128

- Questions or Comments?

  https://github.com/ko-ichi-h/khcoder/issues