A Two-Step Approach to Quantitative Content Analysis : KH Coder Tutorial Using *Anne of Green Gables* (Part II)

HIGUCHI Koichi [i]

Abstract : This article introduces a two-step approach to performing quantitative content analysis of text data. First, an outline of the approach is briefly described. Second, the procedure of using the approach to analyze the novel *Anne of Green Gables* is described as a tutorial. Third, the features of the approach are discussed with reference to the results of the analysis.

The tutorial section of this article allows readers to simulate the same analysis on their own personal computers. We use free software and most of the necessary operations are illustrated in figures. The subject of the analysis is the popular novel *Anne of Green Gables*. It is pointed out that the heroine Anne's foster mother Marilla plays an essential role in the novel and that Marilla is more important than Anne's best friend Diana, and Gilbert with whom Anne has a faint romance. In the analysis of the tutorial, we examine whether the quantitative analysis based on the two-step approach also illustrates the importance of Marilla.

The second half of this article is published here. The first half has been published in Volume 52, Issue 3 of this bulletin.

Keywords : quantitative content analysis, KH Coder, Anne of Green Gables, tutorial

5 Step 2: Focusing on Marilla

5.1 Composing Coding Rules

In Step 1, tables and figures are created in a way less susceptible to the influence of the user's prejudices and preconceptions. In Step 2, however, which is described in this section, the user's viewpoint will be utilized and reflected in the analysis. Nevertheless, while reflecting the user's point of view and interpretation, the process of analysis should be kept open to verification and criticism by third parties. In other words, the user's point of view should be reflected not implicitly but explicitly in the analysis. Composing coding rules is a specific method for performing analysis in that way. A series of such coding rules is sometimes called a "dictionary" for coding.

After coding, the user can count the appearance of the concept or category that he/she focuses on instead of that of each word. For example, Gilbert, a character in *Anne of Green Gables*, is sometimes referred to as "Gilbert" and sometimes as "Gil". The user can count both names as an appearance of a concept "Character Name Gilbert" by composing the following coding rule:

```
*Character_Name_Gilbert
Gilbert or Gil
```

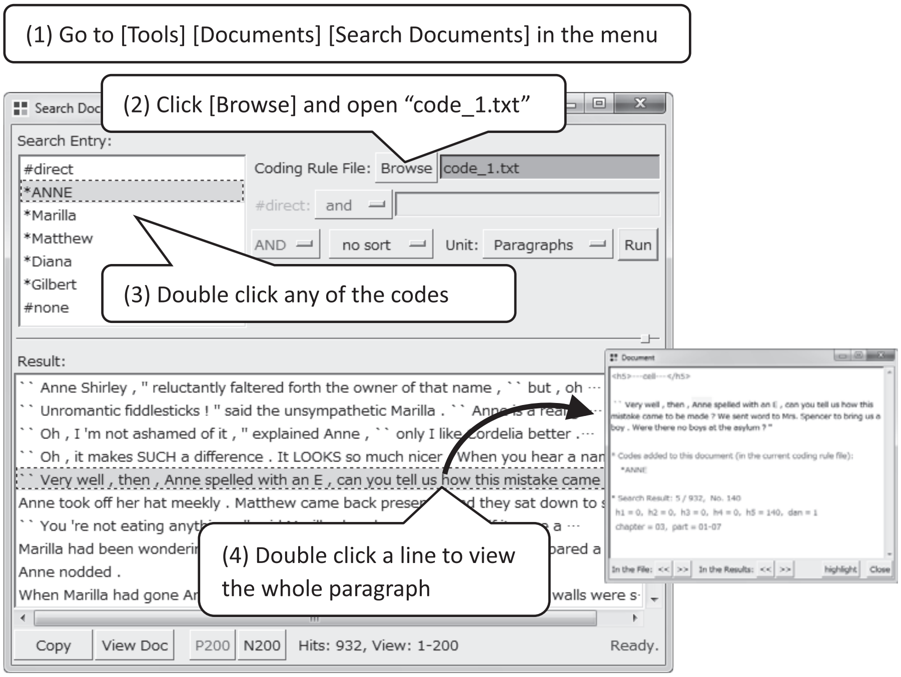
[i] Associate Professor, Faculty of Social Sciences, Ritsumeikan University

Once such a coding rule is entered into KH Coder, not only the documents containing “Gilbert” but also those containing “Gil” are assigned the code “*Character_Name_Gilbert”, and then you can count the appearance of the code. You can compose as many rules as you need and count multiple concepts. If one document satisfies the criteria for multiple coding rules, multiple codes will be attached to the document. Coding of KH Coder is based on the idea of extracting concepts from a document rather than classifying a document into a single category. This concept considers that a document can contain not only one but also multiple concepts.

There are several points to be noted when composing coding rules. First, words not contained in the data and those not extracted as words by KH Coder cannot be counted even if they are designated. Therefore, users are recommended to check whether a word appears in the data by referring to the word list (Section 4.2) and/or by using the word search function (Go to [Tools] [Words] [Search] in the menu) before using the word in a coding rule.

Second, users should confirm what document a code is actually assigned when creating the coding rule. The document search function presented in Figure 12 is a useful means of confirmation. Procedure (2) in Figure 12 shows that the text file attached to the tutorial named “code_1.txt” has been opened. This text file contains coding rules, such as “*ANNE”, “*Marilla”, and “*Matthew”. In procedure (3) in Figure 12, users can double-click any of the codes to retrieve the documents given that code. Here users can also double-click “#none” to retrieve the documents given no codes.

Third, as a general rule, coding rules should be made public to allow third parties to verify whether the researcher has extracted the concepts from the data in a reasonable manner. Even if there is insufficient space to list all the coding rules, users should present some of the main words in each coding rule, and disclose all the coding rules if requested after the research is published.



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

(2) Click [Browse] and open “code_1.txt”

(3) Double click any of the codes

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

Coding Rule File: code_1.txt

#direct: and

AND | no sort | Unit: Paragraphs

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

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 nar
`` 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

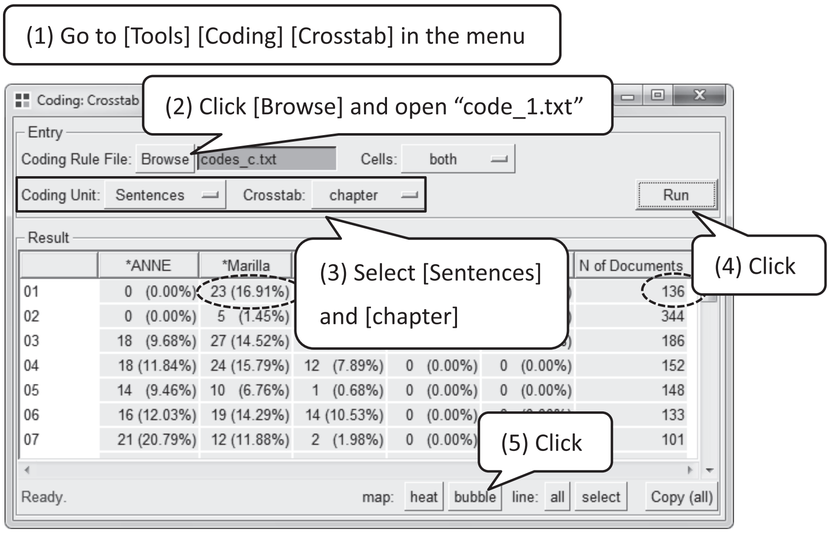
```
<h5>---cell---</h5>

`` 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
```

Figure 12: Retrieve documents assigned a specific code

5.2 Characters in Each Chapter

In this section, we count how many times Marilla appears in each chapter by composing coding rules. Strictly speaking, we count how many sentences contain the name “Marilla” in each chapter. In section 4.4, where we divided the story into four parts, we found that Marilla appears evenly throughout the parts. We then ask, how does she appear in each of the 38 chapters? Let us compare the presence of Marilla with that of other characters chapter-by-chapter in greater detail.

Figure 13 shows the procedure for cross tabulating the results of coding with chapter number. As “Sentences” is selected as the “Coding Unit” in procedure (3) in Figure 13, each sentence is judged if it meets criteria of coding rules. Thus, sentences containing Marilla and other character names are counted. For example, the result pane in Figure 13 shows that 23 sentences contain Marilla in Chapter 01, which is 16.91% of a total of 136 sentences in the chapter. However, it is difficult to read such numbers for all the 38 chapters, so we can create a bubble plot as shown in Figure 14 by procedure (5) in Figure 13.



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

(2) Click [Browse] and open “code_1.txt”

(3) Select [Sentences] and [chapter]

(4) Click

(5) Click

Coding Rule File: codes_c.txt

Cells: both

Coding Unit: Sentences

Crosstab: chapter

	*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%)		133
07	21 (20.79%)	12 (11.88%)	2 (1.98%)	0 (0.00%)		101

Ready.

Figure 13: Cross tabulation of coding results and generation of bubble plots

Figure 14 visualizes the percentage of the sentences containing the name of each character for each chapter. The area of each square is in proportion to the percentage. Furthermore, the color density indicates the degree of difference when compared to other chapters.

Figure 14 shows that Marilla is literally present throughout the entire story. Marilla appears in far more chapters than any other character except for the heroine Anne and almost as widely as Anne. Marilla is the only character who appears throughout the story except Anne, which also suggests the importance of her role.

Chapter 35 is the only chapter where Marilla does not appear. Instead of Marilla, Gilbert appears frequently. In this chapter, Anne is lodging in a town to study at Queen’s Academy and academically competes with her rival Gilbert. Meanwhile, Marilla stays in Avonlea village and does not appear in this chapter. Although they were physically apart, there was an emotional reunion in the following Chapter 36. Anne won a scholarship for being an outstanding student and rejoiced saying “Oh, won’t Matthew and Marilla be pleased!” Also Marilla “smiled affectionately at her girl” and expressed her love to Anne who

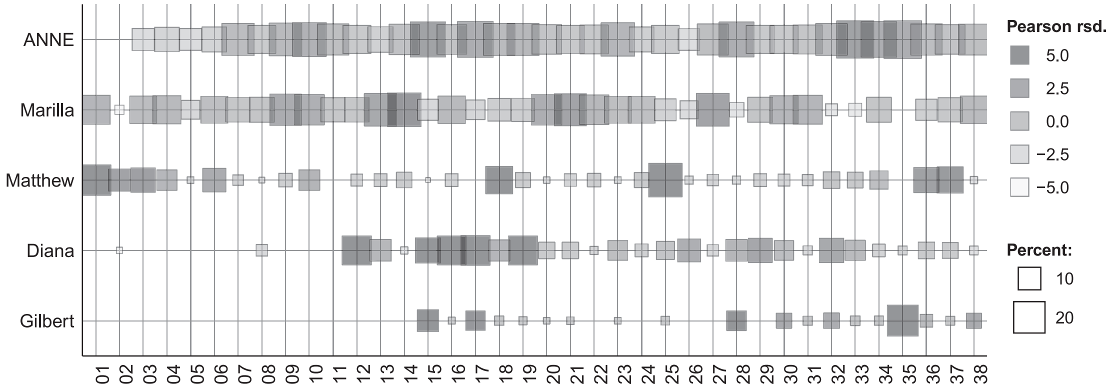
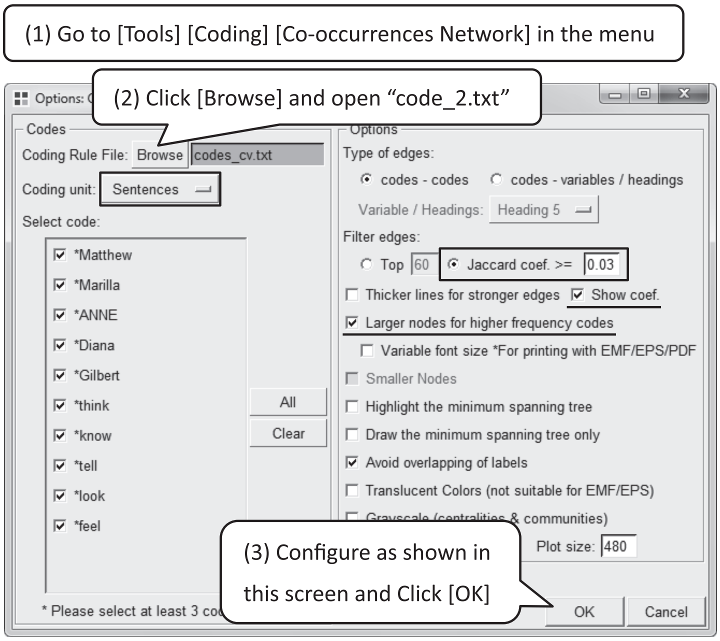


Figure 14: Appearance rate of main characters on a chapter basis

returned to Avonlea. We can speculate that the fact that Marilla and Anne were apart from each other in a chapter helps further deepen the bond between the two in the following chapters.

5.3 Co-occurrence of Characters and Verbs

Next, let us advance our analysis to the detail of the role played by Marilla. By focusing on the co-occurrence of the characters and verbs, we explore what the main characters including Marilla do in the story. Here, we focus on five communication-related verbs: “think”, “know”, “tell”, “look”, and “feel” among the verbs listed in Table 1. Of course “say” in Table 1 is also related to communication, but it co-occurs with all the main characters, and therefore does not manifest the characteristics of individual characters. So, it is excluded from the scope of this analysis[7].



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

(2) Click [Browse] and open “code_2.txt”

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

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

Figure 15: Create a co-occurrence network of codes

A co-occurrence network is created including the main characters listed in Figure 14 and the five verbs mentioned above. Figure 15 shows the procedure of creating it. After the co-occurrence network is saved by KH Coder, it is then edited using Adobe Illustrator to make it easier to read. Here, characters are indicated by gray circles, and verbs, by white circles. Also, co-occurrences of characters are indicated by solid lines, and others, by dashed lines. The result is shown in Figure 16. The numbers in the figure are Jaccard indices, which represent the degree of co-occurrence. The larger the value, the stronger the degree of co-occurrence.

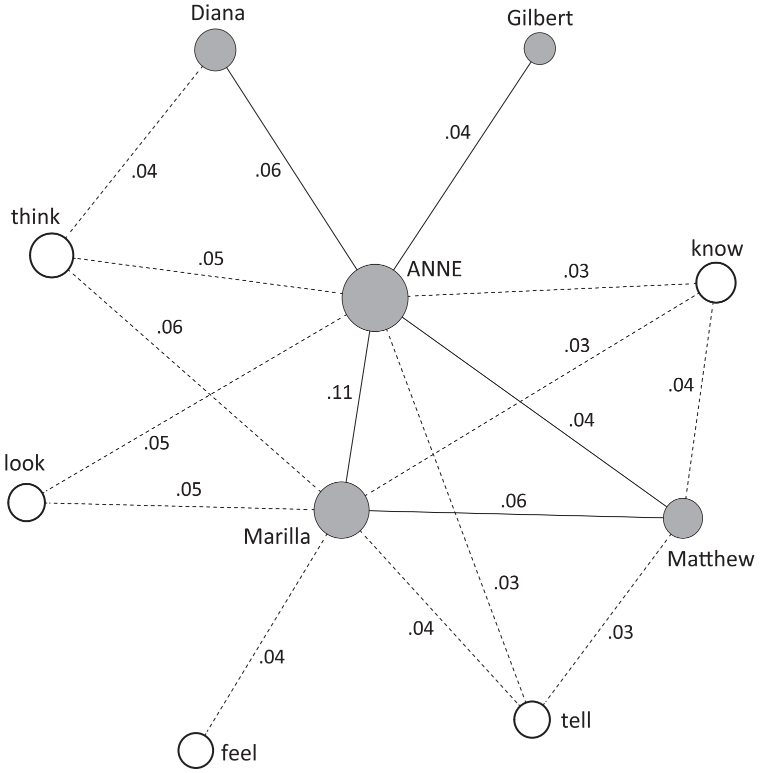


Figure 16: Co-occurrence network of main characters and verbs

Focusing on the links between characters, Figure 16 shows that Gilbert and Diana, who are around the same age as Anne, are independently connected to Anne. In contrast, Anne, Marilla, and Matthew form a triad. This is probably because Anne, her foster mother Marilla, and foster father Matthew are depicted as a group of people closely connected to each other, i.e., a family.

Next, focusing on the verbs co-occurring with Marilla, we observe in Figure 16 that “feel” strongly co-occurs only with Marilla. This means that the verb “feel” is more strongly related to Marilla than to the heroine Anne. This result may be a surprise for many readers, but actually, it does not necessarily mean that Marilla’s feelings are given more importance than Anne’s. Of course there are many descriptions of Marilla’s feelings, but the reason why “feel” strongly co-occurs with Marilla is also because “feel” is often contained in Anne’s words together with “Marilla”, such as “I do feel dreadfully sad, Marilla” (Chapter 21). In this sentence, “feel” does not co-occur with Anne, but only co-occurs with Marilla. The scenes where

Anne expresses her feelings to Marilla are often described in addition to Marilla's own feelings.

Figure 16 also shows that the verb "look" strongly co-occurs only with Marilla and Anne. Then, searching the sentences containing "Marilla", "Anne", and "look", we found that there are many descriptions of how Marilla and Anne look at each other. For example, one of the reasons Marilla decided to adopt Anne was because Marilla looked at, and was moved by, Anne's facial expression.

> Marilla *looked* at Anne and softened at sight of the child's pale face with its look of mute misery—the misery of a helpless little creature who finds itself once more caught in the trap from which it had escaped. (Chapter 6)

Afterward, in the latter half of the story, there is a scene where Marilla gently looks at Anne.

> Marilla *looked* at her with a tenderness that would never have been suffered to reveal itself in any clearer light than that soft mingling of fireshine and shadow. (Chapter 30)

Also, Anne gives a gentle look to Marilla suffering from headache.

> Anne *looked* at her with eyes limpid with sympathy. (Chapter 20)

Anne expresses her feeling of gratitude to Marilla not only by words but also with her eyes, such as "*looked* up earnestly into her face" (Chapter 30).

Thus, Marilla and Anne exchange their feelings not only by words, but also with their eyes, meaning that a close and intimate relationship is depicted between the two.

5.4 Change of Words Co-occurring with Marilla

In the previous sections, we have compared Marilla with other main characters, but in this section, we finally focus on Marilla herself. We make a list of the words co-occurring with Marilla for each of four parts of the story, so that we will be able to explore how Marilla is described and how the description changes as the story progresses.

Following the procedure shown in Figure 17, we can create a list of the words strongly co-occurring with Marilla in part "01–07". To search the words co-occurring with Marilla in the following part "08–19", repeat procedure (3) in Figure 17 and then click [*08–19] instead of [*01–07] in procedure (4). Do the same for all the four parts and list the top 10 words for each part to create Table 2. In Table 2, the cells containing the words the author particularly noted are hatched.

Table 2 shows that the word most characteristic of Marilla in part "01–07" is "Matthew", meaning that there are frequent interactions between Marilla and Matthew. Also, not "Anne" but "child" is listed for the first part, suggesting that Anne is not called by her own name, but likely to be treated as a nameless "child" by Marilla in this part. Additionally, Marilla is not used to treating a "child", and therefore an "uncomfortable" situation occurs as follows:

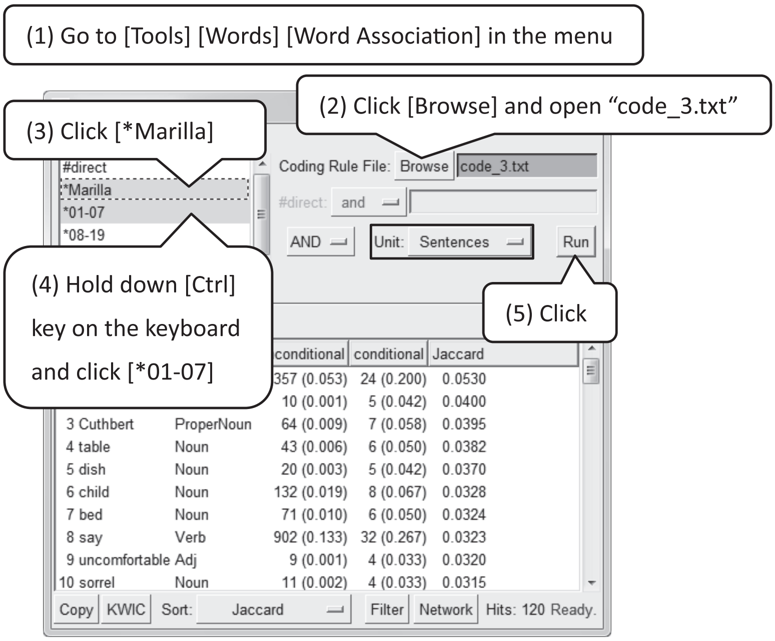
> Marilla really did not know how to talk to the child, and her *uncomfortable* ignorance made her crisp and curt when she did not mean to be. (Chapter 04)

In the first part, Marilla is not necessarily kind to the child, but is rather a person of few words.

However, for the following parts "08–19" and "20–28", "say" and "Anne" are listed as the words characteristic of Marilla (Table 2), showing that the "child" is upgraded to "Anne" and implying that it is impossible to bring up a child without "saying" anything.

> "I don't think there is much fear of your dying of grief as long as you can talk, Anne", *said* Marilla unsympathetically. (Chapter 17)

Such a description does not suggest that Marilla indulges Anne, but obviously shows that Marilla's attitude has appreciably changed from "uncomfortable ignorance" in the first part. Along with that, in the part "20–



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

(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

#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)	0.0328
7 bed	Noun	71 (0.010)	6 (0.050)	0.0324
8 say	Verb	902 (0.133)	32 (0.267)	0.0323
9 uncomfortable	Adj	9 (0.001)	4 (0.033)	0.0320
10 sorrel	Noun	11 (0.002)	4 (0.033)	0.0315

Copy KWIC Sort: Jaccard Filter Network Hits: 120 Ready.

Figure 17: Search related words

Table 2: Change of words co-occurring with Marilla

01-07		08-19		20-28		29-36	
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 values are Jaccard indices, which represent the degree of co-occurrence.

28", "feel" is listed as a characteristic word, meaning that the scenes where Anne expresses her feelings to Marilla come to be described as well as Marilla's own feelings (Section 5.3).

In the last part "29-36", Matthew appears again (Table 2). This is probably because Matthew passes away almost at the end of the story. After Matthew's death, Marilla and Anne talk about Matthew. Here, "look" appears for the first time as a characteristic word. As described in Section 5.3, Marilla's and Anne's eyes on each other is depicted by the word "look". For example, in the scene where Anne says that she has decided to quit leaving home and stay with Marilla, who is anxious about her health, there is a description as follows:

> "Not going to Redmond!" Marilla lifted her worn face from her hands and *looked* at Anne. "Why, what do you mean?" (Chapter 38)

Their strong emotions are expressed with their eyes as a proverb says "The eye is the window of the mind".

Thus, Table 2 shows that Marilla, who once was not used to treating children and "crisp and curt", changes her attitude as the story progresses. Marilla gradually changes herself to make a close and intimate relationship, such as exchanging feelings and emotional eye contacts, with Anne. Marilla not only appears frequently, but also plays an essential role of gradually making a deep and rich relationship with Anne in this story.

The above quantitative analysis supports the assertion made by Doody (1997) that the education of Marilla is the central theme of the story (Section 2.1). The most important keywords among those suggested by the quantitative analysis include "child" and "uncomfortable" in early parts, and "feel" and "look" in latter parts. Identifying such keywords through quantitative analysis is considered to be useful for extracting depiction which specifically describes Marilla's change or education.

6 Features of Two-Step Approach

6.1 Advantages of Quantitative Analysis

The primary advantage of performing quantitative analysis as introduced in this article is that it allows for exploring data, in other words, contributes to better understanding of the data.

One aspect of data exploration by quantitative analysis is that we can obtain overviews of entire data. By automatically counting words using a computer, you may notice, for example, that "Marilla appears more frequently than I thought". Overviews of data can be useful by itself and allow us to discover features of data that have not been previously observed. Also, overviews are useful when attempting to analyze the meaning of each word or phrase in detail, since trying to examine the meaning of one sentence in detail makes it difficult to simultaneously view the data as a whole. This issue is compensated by visualizing entire data into a form of a graph or map using quantitative methods. For example, if you obtain an overview of the story flow by performing correspondence analysis (Figure 11), it will be easier to examine a character's role in a specific scene.

In addition to being able to obtain an overview of the data, another important aspect of data exploration is that the data can alert researchers to important sentences that should be read in detail. For example, if you find that sentences containing "Marilla" often contain "feel" too, you may notice something new about the data by searching for sentences containing both "Marilla" and "feel". Also, if the appearance of a specific character suddenly increases or decreases among bubble plots such as shown in Figure 14, there may be some movement in the story. Thus, quantitative analysis suggests which part of the data is considered to be important and which part of the data is to be interpreted in detail by researchers.

The secondary advantage of quantitative analysis is that it improves the reliability of analysis. Actually, this point is inextricably associated with the advantage that it allows for exploring the data as described above. This is because if an overview of the data is presented by quantitative analysis, it allows third parties to check whether the user envisions a biased, selective whole image convenient for his/her hypotheses or theories. Also, you will be able to address such doubts as "How did you choose the sentence to quote or interpret from the data?" or "Didn't you quote only the parts convenient for you?" to some extent. For example, you will be able to explain that you focused parts where a sudden change occurs in the graph or you searched for sentences containing the word “feel” that was characteristic to a character. As described above, you can clearly demonstrate how the conclusion was derived from the data, thus improving the reliability of the analysis. This will help to accumulate research that can withstand comparison and verification.

6.2 Academic Background of Two-Step Approach

The author proposed a two-step approach introduced in this article intending to make the advantages of the quantitative analysis described above easier to be used by more researchers (Higuchi 2004, 2014). This approach is based on the idea of content analysis. In this section, the academic background and philosophy of this approach are described as the closing remark of this article.

Computers have been actively used for content analysis since the 1960s. At that time, there were two conflicting ideas regarding analytical approaches (Stone 1999). The first approach found groups of words that often appear together in the same sentence through statistical analysis, and is now called “Correlational approach”[8]. The other is an analysis method for extracting concepts from data by using coding rules, and is now called “Dictionary-based approach”. The advocates of the Correlational approach argued that the advantage of their approach was that the analytical results are not “contaminated” by the researcher’s theories, hypotheses, or prejudices (Iker & Harway 1969). Meanwhile, the advocates of the Dictionary-based approach thought that coding was essential to achieve the purpose of analysis (Osgood et al. 1957).

At first glance, the two-step approach introduced in this article may seem to be a simple joining of these two approaches. However, in Step 1, there are some differences from the conventional Correlational approach. That is, when the conventional Correlational approach was used in actual research, words to be analyzed were often hand-selected, and words with similar meanings, such as “say” and “talk”, were designated a singular unit. If you perform such manual designations many times, contrary to the above argument by Iker & Harway (1969), the preconceptions of the researcher may be implicitly introduced. Therefore, we decided not to do such manual work at the first stage of the two-step approach introduced in this article. By doing so, you are spared the effort of performing manual designations while improving the reliability of analysis results.

Also, with the conventional Dictionary-based approach, composing coding rules was never an easy task. Composing coding rules used to require time and effort consuming tasks such as examining the entire set of data (Saporta & Sebeok 1959). Also, it was difficult for third parties to judge the appropriateness of the content of the coding rules. However, with the two-step approach introduced in this article, you can compose coding rules referring to the overview of the data that is revealed in Step 1. Also, you only have to write coding rules for the concepts that need to be additionally extracted. Therefore, it is much easier to compose coding rules than with the conventional Dictionary-based approach. Furthermore, by comparing the result of Step 1 analysis with the coding rules, any third party can confirm on which part of the data the researcher has focused. The reliability of analysis is improved also in this respect.

Thus, the two-step approach introduced in this article is to join the existing two approaches while making modifications. With this modification and joining, this approach has made content analysis easier and more reliable.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 26705006.

Notes

7 As long as operated explicitly as above, users are recommended to proactively focus on some words or concepts in Step 2 in this manner.

8 Correlational approach is also called as Statistical Association approach.

References

Danowski, J. A., 1993, “Network Analysis of Message Content”, W. D. Richards Jr. & G. A. Barnett eds., *Progress in Communication Sciences IV*, Norwood, NJ: Ablex, 197-221.

Doody, M. A. 1997, “Introduction”, W. E. Barry, M. A. Doody and M. E. D. Jones eds. *The Annotated Anne of Green Gables*, Oxford University Press, New York, 9-34.

Greenacre, M. J., 2007, *Correspondence Analysis in Practice 2nd ed.*, Boca Raton, FL: Chapman & Hall/CRC.

Higuchi, K., 2004, “Quantitative Analysis of Textual Data: Differentiation and Coordination of Two Approaches”, *Sociological Theory and Methods*, 19(1): 101-15 (Written in Japanese).

Higuchi, K., 2014, *Quantitative Text Analysis for Social Researchers: A Contribution to Content Analysis*, Nakanishiya Publishing: Kyoto, Japan (Written in Japanese).

Higuchi, K., 2016, “A Two-Step Approach to Quantitative Content Analysis: KH Coder Tutorial Using Anne of Green Gables (Part I)”, *Ritsumeikan Social Science Review*, 52(3): 77-91.

Iker, H. P. & N. I. Harway, 1969, “Computer Systems Approach toward the Recognition and Analysis of Content”, G. A. Gerbner, O. R. Holsti, K. Krippendorff, W. J. Paisly & P. J. Stone eds., *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*, New York: Wiley & Sons, 381-486.

Kawabata, Y., 2008, “Surprise of Marilla Cuthbert” Katsura, Y. and Shirai, S. eds. *The world of Masterpieces We Want to Know More 10: Anne of Green Gables*, Minerva: Kyoto, Japan, 109-19 (Written in Japanese).

Matsumoto, Y., 2008, *Journey to the Anne of Green Gables: Hidden Love and Mystery*, NHK Publishing: Tokyo, Japan (Written in Japanese).

Osgood, C. E., 1959, “The Representational Model and Relevant Research Methods,” I. d. S. Pool ed., *Trends in Content Analysis*, Urbana, IL: University of Illinois Press, 33-88.

Osgood, C. E., G. J. Suci & P. H. Tennenbaum, 1957, *The Measurement of Meaning*, Urbana, IL: University of Illinois Press.

Saporta, S. & T. A. Sebeok, 1959, “Linguistic and Content Analysis,” I. d. S. Pool ed., *Trends in Content Analysis*, Urbana, IL: University of Illinois Press, 131-50.

Stone, P. J., 1997, “Thematic Text Analysis: New Agendas for Analyzing Text Content,” C. W. Roberts ed., *Text Analysis for the Social Sciences*, Mahwah, NJ: Lawrence Erlbaum, 35-54.

Yamamoto, S. 2008, *From Anne Shirley to Jane Eyre: Introducing English Literature in University Classes*, University of Tokyo Press: Tokyo Japan (Written in Japanese).

接合アプローチによる量的内容分析の実践（二）
―『赤毛のアン』を用いたKH Coderチュートリアル―

樋口　耕一[i]

本稿では，量的な内容分析を実践するための方法として筆者が提案している「計量テキスト分析」を，新たな分析事例とともに紹介する。計量テキスト分析において，データを分析する具体的な手順にはいくつかのバリエーションがあるが（Higuchi 2014），本稿では特に「接合アプローチ」と呼ばれる手順をとりあげる。第一に，このアプローチと，その実現のために筆者が開発・公開しているフリーソフトウェアKH Coderについて概要を手短に紹介する。第二に，このアプローチにもとづいて小説『赤毛のアン』を分析する手順を，読者が自分のPCで同じ分析を行えるチュートリアルの形で記述する。第三に，分析の結果を踏まえて，本アプローチの特徴について議論する。

本稿で紹介する接合アプローチとは，従来の内容分析で利用されてきた２つのアプローチを接合したものである。従来の内容分析では，テキスト型データを計量的に分析するためにCorrelationalアプローチかDictionary-basedアプローチを用いることが多かった。Correlationalアプローチはクラスター分析のような統計手法を用い，頻繁に同じ文書の中にあらわれる言葉のグループを見つけだすといった方法で，データ中の主題を探索するアプローチである。このアプローチはStatistical Associationアプローチと呼ばれることもある。それに対してDictionary-basedアプローチでは，統計手法ではなく，分析者自身の指定した基準にそって言葉や文書を分類し，計量的な分析を行なう。これら２つは考え方が大きく異なるアプローチでありながら，実際の分析においては混同されやすい部分もあった。そこで混同されやすい部分を峻別した上で，これら２つを接合したものが，本稿で紹介する接合アプローチである。

本稿のチュートリアルでは，この接合アプローチを用いて，小説『赤毛のアン』の原文を分析する。小説『赤毛のアン』では，主人公である孤児のアンが，マシューとマリラの兄弟に引き取られ，成長していく様子が描かれている。この物語においては養母マリラの果たした役割が非常に大きいという指摘がある。親友のダイアナや，アンとの淡いロマンスが描かれるギルバートよりも，マリラの方が中心的であったという。また『赤毛のアン』は，マリラが子供を愛することを学び，それによって自分自身も幸せになっていくという，大人の成熟と生き直しの物語であると指摘されている。本稿の分析では，こうしたマリラの重要性を，計量的分析からも読み取ることができるのかどうかを確認する。

なお本稿の後半をここに掲載する。前半については本誌の52巻３号に掲載している。

キーワード：量的内容分析，KH Coder，赤毛のアン，チュートリアル，計量テキスト分析

[i] 立命館大学産業社会学部准教授