

Popularization of Classical Economics:
The text-mining Analysis of David Ricardo,
James Mill, and Harriet Martineau

Hiroyuki Shimodairaⁱ

Shinji Fukudaⁱⁱ

Research Group of Economics and Management
No. 2014-E01
2014.12

ⁱ Faculty of Literature and Social Sciences, Yamagata University, E-mail: shimo [at] human.yamagata-u.ac.jp

ⁱⁱ Faculty of Humanities, Hirosaki University, E-mail: shinji [at] cc.hirosaki-u.ac.jp

**Popularization of Classical Economics:
The text-mining Analysis of David Ricardo, James Mill, and Harriet Martineau**

Hiroyuki Shimodaira & Shinji Fukuda

1 Introduction

1.1 The background of the study

Ricardo's *Principles of Political Economy* (1st ed., 1817) established the theoretical foundations of classical economics and determined the direction of the later history of economics. However, Ricardo's *Principles* was not necessarily accepted directly into society at large. It is well established that certain individuals --- including James Mill and Harriet Martineau, who was not a top class theorist in the field --- played an important role in the dissemination of Ricardian economics. In other words, classical economics spread throughout over the general public by way of popular textbooks or enlightenment works such as Mill's *Elements of Political Economy* (1st ed., 1821) and Martineau's *Illustrations of Political Economy* (9 vols., 1st ed., 1832-34), as well as through technical books such as Ricardo's *Principles*. Therefore, due to the mechanisms of diffusion, not only are there problems regarding the technical content of economics, but these problems are magnified and transformed into popular beliefs. Certainly, this has been widespread issue found in the overall history of economics.¹

If we consider such circumstances, we must examine more popular literature as well as specialized discourse in order to study the general process of dissemination of economics. It is expected that such literature will amount to enormous quantities. Even the principal technical books throughout the history of economic thought are greatly numbered, and it is difficult to even elucidate those contents. Therefore, in order to examine the enormous body of more popular literature in addition to detailed scholarly works, a method distinct from the conventional qualitative study of economic literature is needed to fulfill this investigation of the history of economic thought. In this paper, we propose text-mining analysis as a means to investigate this question. Text-mining analysis is a technique that enables us to statistically process the text data of various literatures and to examine such literature quantitatively. The following section will outline this process in further detail. In this paper, we will investigate whether such analysis is effective in the study of the history of economic thought while focusing on the process of the dissemination of classical economics.

In addition, this paper is associated with the the project, 'A proof study of the process of reception and spread of economic thought: How did people receive economics?'² This project is an attempt to qualitatively and quantitatively categorise the process by which economic theory and thought have disseminated from economic specialists to non-specialists, that is, the general public, through the last 200 years beginning with the formation of classical economics, while focusing the

research on the history of economic thought in England. Our primary approach attempts to clarify the process underwent by the theories and thoughts in the representative literatures of economics (Adam Smith's *Wealth of Nations*, Karl Marx's *Capital*, John Maynard Keynes' *General Theory*, among others) as they were accepted into the general public through more popular criticisms, primers, textbooks, commentary books, popularity books, poetry, fables, novels, brochures, and handbills, among others.

1.2 The task of this paper

The task of this paper is to examine the process of dissemination of classical economics through a text-mining analysis of major works, as a part of the project mentioned above. In particular, we will clarify some aspects of the process of popularization of Ricardian economics. For this purpose, we will examine Ricardo's *Principles* (3rd ed.), Mill's *Elements* (3rd ed.), and Martineau's *Illustrations*, No.25 (3rd ed.), and compare them, using the text-mining analysis procedure. Martineau is known for her important role in the dissemination of classical economics; most notably, Ricardo's arguments in his *Principles* were spread into the general public by Martineau's works, *Illustrations*. What kind of transformation did Ricardian theory undergo at this time? We seek to clarify that process while referring to Mill's *Elements*. With these three formative works, we aim to investigate the validity of using a text-mining analysis to study of the process by which economics is disseminated.

Because Martineau assumed that her audience in *Illustrations* would be the general public, in this paper we set the following two hypotheses;

- (1) Martineau removes technical terms, and uses a many popular words easily understood by the general public.
- (2) In contrast with Ricardo and Mill, Martineau contextualizes the discourse with real world policies and maintains a broader worldview in the background rather than simply relying on basic theory analysis.

In the following sections, we investigate whether such hypotheses are recognized as appropriate, and whether they are supported by the text-mining analysis.

2 The outline of the text-mining analysis and the text data

2.1 The study of history of economic thought and the text-mining analysis

The text-mining analysis is 'a generic name of the various technique to analyse text data quantitatively with the calculator and to extract a useful law' (Matsumura & Miura 2009: 1). Originally, 'mining' refers to the act of 'digging out coal'. Therefore, the text-mining analysis includes two basic characteristics. The first is to extract a law or knowledge hidden in a large quantity of information. The second is to carefully select the meaningful, important, and useful laws from the various extracted information. At the first stage, we use mathematic processing

(quantitative analysis), mainly employing a statistics processing software. However, at the second stage, we must make a judgment in order to select what is not self-evident, but what is meaningful, in regards to laws, interpretations, and knowledge; that is, we implement a qualitative analysis. In other words, the text-mining analysis provides a quantitative perspective to supplement the qualitative analysis.

Therefore, the text-mining analysis creates the possibility to connect with a conventional method of the study of the history of economic thought. In the text-mining analysis, we first collect the digitized data of the literature, then adjust the data for statistical processing, process the adjusted data, and obtain a result. However, the text-mining analysis is not concluded with this technical processing only. We must provide our 'interpretation' at every stages; that is, when regarding the hypothesis, the meaning of the collected data, the naming of categories, the significance of the results and their limits, and when considering the importance placed upon the process in general. For this reason, the text-mining analysis absolutely requires the knowledge of previous studies of the history of economic thought, which have been accumulated throughout many years. In applying this prior knowledge, we may discover new knowledge, for example, the identification of the writer of an anonymously authored work, clarification of changes in different editions of a book, and the transformation of popularized verbal explanation that has been previously unnoticed. These discoveries are made possible through statistically processing a large quantity of the text data in the literature.

In other words, we expect to find more general and more persuasive knowledge by applying the quantitative statistical analysis to a large amount of text data than that provided by individual example report using a qualitative analysis to examine a smaller body of work. In this sense, the text-mining analysis is completely compatible with the traditional methods used to study the history of economic thought; it may establish itself as a new method to study of the history of economic thought that further develops the field, bringing about a more heuristic knowledge of the subject.³

2.2 The outline of the text-mining analysis

In this paper, we will compare and contrast Ricardo's *Principles*, Mill's *Elements*, and Martineau's *Illustrations*, No.25, using frequency analysis and co-occurrence network analysis, which form the representative techniques of the text-mining analysis. Before the analysis, we outline the process of these techniques in detail. In this study, we use the KH Coder⁴ as the text-mining software in these analyses.

(1) The frequency analysis

The frequency analysis examines the frequency of certain words in the analysed literature. In this paper, we extract nouns that appear to be highly frequent in the literature using the KH Coder. We focus our attention on nouns in this paper based on a proposition of the categorization study in cognitive psychology; that is, 'it is nothing but in noun that a certain idea, concept, and thought appear most symbolically' (Kida 2008: 151). At its core, the frequency analysis reflects a basic

logic that ‘a word of higher frequency of appearance is higher in importance’ and ‘a word of lower frequency is lower in importance’; these assumptions allow us to characterize the selected literature.

(2) The co-occurrence network analysis

The co-occurrence network analysis seeks to examine the relationship between various words in the literature. When we are not able to sufficiently clarify the characteristics of the literature by using the frequency analysis, we may provide further clarification through an examination of the relationship between the words in the literature by using the co-occurrence network analysis. Using the KH Coder, we will be able to elucidate a relationship between the words, by connecting words that have similar patterns of appearance with lines. This indicates a strong degree of co-occurrence, and allows us to further visualise these relationships by drawing a figure showing the network of those lines.

Regarding the relative strength of the co-occurrence, we can find the value of the Jaccard coefficient that is calculated with respect to all combinations of the words in the literature. The Jaccard coefficient takes a value between 0 and 1; as the frequency of two words appearing (co-occurring) in the same sentence increases, the value becomes closer to 1. If the two words do not appear (do not co-occur) in the same sentence, the value is closer to 0. The Jaccard coefficient (J) of the co-occurrence between any word X and any word Y is calculated in the following equation:

$$J=A/(A+B+C)$$

Here, A expresses the number of the sentences where both word X and word Y appear at the same time, B expresses the number of the sentence where only word X appears, and C expresses the number of the sentence where only word Y appears.

The KH Coder automatically calculates the value of the Jaccard coefficient, and draws a figure displaying the network between the words. Concerning the network figure form, we can choose one of two patterns of figures: (1) a network based on the ‘centricity’ of the words, and (2) a network introduced by the ‘sub-graph detection’. These figures are used in order to identify commonalities among the various literatures studied.⁵

- (1) Centricity analysis: With this analysis, the network figure shows the extent which each word plays a central role in the literature. Each word is distinguished by colour; their degree of the centricity is shown in order of red, pink, white, light blue, with red showing the greatest.
- (2) Sub-graph detection: This analysis reveals words that are strongly tied to one another in the literature, and the network figure displays groups of these words. Each group is distinguished by colour: green, yellow, blue, red and so on.⁶ A white colour indicates that it is an independent word that does not form a group with other words. A solid line connects words included in the same group; words included in different groups are connected by a broken line.

In addition, words with stronger co-occurrence are connected by a thicker line in both networks.

figures.

2.3 The text data of the literature

For our text-mining analysis, we use the text data from the sources listed below:⁷

Ricardo, D. 1821, *The Principles of Political Economy, and Taxation*, 3rd ed., London: John Murray.

Mill, J. 1844, *Elements of Political Economy*, 3rd ed., London: Henry G. Bohn.

Martineau, H. 1834, *Illustrations of Political Economy*, 3rd ed., vol.9, London: Charles Fox.

We digitize each of the text data of these literatures according to the following procedure:

- (1) Scan the text data and process it with the Optical Character Reader (OCR).
- (2) Shape the data by manually correcting a mistaken data and removing a meaningless signs using software such as Microsoft Word.
- (3) Give these data a tag and save it as a text file.

After this process, we analyse the digitized data using the KH Coder.

3 The results of the analysis

3.1 The frequency analysis

We first applied the frequency analysis to Ricardo's *Principles*, Mill's *Elements*, and Martineau's *Illustrations*, No.25, and extracted the nouns that rank from 1 to 30 in frequency of appearance in each document; results are displayed in Table 1.⁸ Thus, we are able to clarify the following points in comparing the highest ranking nouns in each document:

- (1) The ten highest-ranking nouns (indicated in red) in Ricardo's *Principles* are included in all of the 30 highest-ranking nouns in Mill's *Elements*, but only five nouns among them are found in the top 30 ranking nouns in Martineau's *Illustrations*, No.25.
- (2) Among the nouns ranked 11-20 in Ricardo's *Principles*, nine of these are included in the 30 highest-ranking nouns in Mill's *Elements*, but only six nouns are found among those in Martineau's *Illustrations*, No.25.
- (3) Among the nouns ranked 21-30 in Ricardo's *Principles*, only two of these are included in the 30 highest-ranking nouns in Mill's *Elements* or Martineau's *Illustrations*, No.25.
- (4) In the 30 highest-ranking nouns in Mill's *Elements*, nine nouns are found that do not appear in the 30 ranking of nouns in Ricardo's *Principles*. Furthermore, only three words of these words appear in the 30 highest-ranking nouns in of Martineau's *Illustrations*, No.25.
- (5) In the 30 highest-ranking nouns in Martineau's *Illustrations*, No.25, 14 of these (indicated in blue) are not found in the 30 highest-ranking nouns in either Ricardo's *Principles* or Mill's

Elements.

3.2 The co-occurrence network analysis (1): the centrality analysis

Next, we applied the co-occurrence network analysis to the nouns in Ricardo's *Principles*, Mill's *Elements*, and Martineau's *Illustrations*, No.25, and drew figures of the network based on their 'centricity'; these are expressed in Figures 1.1-1.3.⁹ We then made the following observations based on a comparison of these figures:

- (1) In Ricardo's *Principles*, 'price' and 'profit' play the most central role, with 'produce' and 'labor' being of secondary importance.
- (2) In Mill's *Elements*, 'labor' and 'commodity' play the most central role, with 'wages', 'capital', and 'money' having secondary importance.
- (3) In Martineau's *Illustrations*, No.25, 'capital' and 'society' play the most central role.
- (4) The most centrally important nouns among the three texts are all different from each other.

3.3 The co-occurrence network analysis (2): the sub-graph detection

Lastly, we drew the figures of the network by applying the 'sub-graph detection' to Ricardo's *Principles*, Mill's *Elements*, and Martineau's *Illustrations*, No.25; these are presented in Figures 2.1-2.3. A comparison of these figures enabled the following observations:

- (1) In Ricardo's *Principles*, the first group (in green) includes 'money', 'value' and 'corn', et al., the second group (in yellow) includes 'land', 'produce', and 'rent', et al., and the others are found.
- (2) In Mill's *Elements*, the first group (in green) includes 'proportion', 'case', and 'labourer' et al., the second group (in yellow) includes 'price', 'commodity', and 'quantity', et al., and the others are found.
- (3) In Martineau's *Illustrations*, No.25, the first group (in green) includes 'labourer', 'capitalist', and 'wages', et al., the second group (in red) includes 'country', 'commodity', and 'money', et al., the third group (in blue) includes 'food', 'supply', and 'demand', et al., and the others are found. In addition, the groups are more decentralized and further divided in comparison with the previous two works.
- (4) In the first group of Ricardo's *Principles*, the second group of Mill's *Elements*, and the second group of Martineau's *Illustrations*, No.25, the components that are common among two or three of the texts, such as 'money' and 'commodity', among others, are found, but their construction within the chart differ greatly from one another.
- (5) In the second group of Ricardo's *Principles*, the first group of Mill's *Elements*, and the first group of Martineau's *Illustrations*, No.25, words such as 'produce' and 'labourer' are found in two of the groups, but these do not appear among all three network.

4 Considerations

4.1 The characteristics of Ricardo's *Principles*

In his *Principles*, Ricardo examined theoretical and policy issues in relation to value, distribution, growth, trade, tax, and money, based on his labour theory of value.

First, according to the frequency analysis in Table 1, the ten highest-ranking nouns in Ricardo's *Principles* are 'price', 'value', 'commodity', 'labor', 'capital', 'quantity', 'profit', 'tax', 'corn', and 'money'. This result reflects the arguments made in Ricardo's *Principles*. It is important to note that the frequency of highest ranking noun, 'price', and the second-highest ranking noun, 'value', greatly surpasses the others; this suggests that Ricardo placed a greater importance on the issue of determining price or value. In addition, the concepts such as 'rent' and 'wages', which are of great importance to Ricardo, are found in the nouns ranking 11-20 in the table. Although the focus seems to dim slightly, important concepts such as 'gold' and 'trade' are also found in the ranking of nouns from 21-30.

Next, based on the centrality analysis of the co-occurrence of words in Figure 1.1, we observe that 'price' and 'profit' play the most central role in Ricardo's *Principles*, and 'produce' and 'labor' play a slightly less crucial role. This result is reminiscent of Ricardo's arguments in his *Principles* concerning the relation between the decision of price and the falling rate of profits.

Moreover, according to Figure 2.1 of the sub-graph detection of the co-occurrence network analysis, the first group of nouns (in green) in Ricardo's *Principles* consists of 'money', 'value', 'corn', 'commodity', 'quantity', 'production', 'labor', and 'price'. This group primarily reflects Ricardo's arguments on the labour theory of value and the changing of corn prices in his *Principles*. In addition, the second group (in yellow) consists of 'land', 'produce', 'rent', 'profit', 'tax', and 'landlord'. This group mainly reflects Ricardo's arguments on the distribution and taxation in his *Principles*.

In this way, the results of the text-mining analysis appropriately, although roughly, illustrate defining characteristics of Ricardo's *Principles*.

4.2 The characteristics of Mill's *Elements*

Mill, known to be a friend of Ricardo, pressed Ricardo to write his *Principles*. Mill subsequently organized his own notes on the subject in order to teach his son, John Stuart Mill, the elements of Ricardo's *Principles*, and wrote his *Elements* as an introduction to Ricardian economics. In his *Elements*, Mill explains the principles of Ricardian economics while dividing them into the four branches: production, distribution, exchange, and consumption.

According to the frequency analysis in Table 1, the ten highest-ranking nouns in Ricardo's *Principles* are all included in 30 highest-ranking nouns in Mill's *Elements*, and nine nouns among those ranked 11-20 in Ricardo's *Principles* are included in 30 highest-ranking nouns in Mill's *Elements*. These results reflect the fact that Mill's *Elements* is relatively faithful to Ricardo's *Principles*. However, the highest ranking word in Ricardo's *Principles*, 'price', drops down to rank

16th in Mill's *Elements*; similarly, the ninth-ranking word, 'corn', drops down the rank 23rd. Additionally, only two nouns among those ranked 21-30 in Ricardo's *Principles* are found in the highest 30 in Mill's *Elements*, and concepts such as 'gold' and 'trade', which were deemed as important to Ricardo, are not included at all. These results suggest that the central arguments in Ricardo's *Principles* were slightly diluted in Mill's *Elements*.

Next, according to Figure 1.2 displaying the centrality analysis of the co-occurrence network, 'labor' and 'commodity' play the most central role in Mill's *Elements*, and 'wages', 'capital', and 'money' play a subordinate, although still important, role. This result suggests that the central arguments in Ricardo's *Principles* do not necessarily occupy an identical position in Mill's *Elements*.

Moreover, according to the sub-graph detection of the co-occurrence network analysis in Figure 2.2, the first group of nouns (in green) in Mill's *Elements* consists of 'proportion', 'case', 'labourer', 'wages', 'profit', 'labor', 'capital', 'portion', 'rent', 'produce', and 'land'. This group includes the components that are common to the second group of Ricardo's *Principles* (listed here in italics), so it shows correspondence, to some extent, with the latter. In addition, the second group of nouns (in yellow) in Mill's *Elements* consists of 'price', 'commodity', 'country', 'quantity', 'money', 'value', and 'metal'. This group includes the components (in italics) that are common to the first group of Ricardo's *Principles*, again showing a degree of correspondence to the latter. These results further demonstrate that Mill's *Elements* follows Ricardo's *Principles* fairly consistently. However, in Mill's *Elements*, a small group of nouns (in orange, red, purple, and blue), which are not observed in Ricardo's *Principles*, is found. It is particularly interesting to note that the appearance of the small group consisting of 'supply' and 'demand' is found. These results suggest that the construction of economic arguments is more decentralized in Mill's *Elements* than in Ricardo's *Principles*.

In this way, the results of the text-mining analysis show that Mill's *Elements* follows Ricardo's *Principles* in a relatively faithful manner. At the same time, the construction of the arguments relating to the labour theory of value and the decision of corn prices, elements inherent to Ricardo's *Principles*, is somewhat diluted in Mill's *Elements*.

4.3 The characteristics of Martineau's *Illustrations*, No.25

Martineau played a particularly important role in the dissemination of classical economics. Martineau drew the various morals derived from the principles of classical economics and rewrote them in a fable style so that the general public would easily understand them. This is presented in her *Illustrations*, which consists of nine volumes containing 25 fables. In the fable, No.25, Martineau showcases the principles of economics that form the theoretical bases for fables, Nos.1-24. Here, Martineau relies on the construction of the four branches of economics from Mill's *Elements*.¹⁰

First, according to Table 1 showing the frequency analysis, only five nouns among the ten highest-ranking nouns in Ricardo's *Principles* (specifically, 'value', 'commodity', 'labor', 'capital', and 'money') are included in the top 30 in Martineau's *Illustrations*, and the other five

nouns ('price', 'quantity', 'profit', 'tax' and 'corn') are not included; of particular note is the absence of the highest-ranking nouns in Ricardo's *Principles*, 'price'. Nouns which do appear in both Ricardo's *Principles* and Martineau's *Illustrations*, No.25, see a large decrease in rank in the latter text; for example, 'value', which is the 2nd most frequent noun in Ricardo's *Principles*, drops down the rank the 21st in Martineau's *Illustrations*, No.25. In place of these nouns, 'labor' becomes the highest-ranking noun, followed by 'capital'. Moreover, only six nouns among those ranked 11-20 in Ricardo's *Principles* are included in the top 30 most frequent nouns in Martineau's *Illustrations*, No.25, and only two nouns among those ranked 21-30 in Ricardo's *Principles* appear in the ranking of *Illustrations*, No.25. On the other hand, 14 nouns not included in 30 most frequent nouns in either Ricardo's *Principles* or Mill's *Elements* are found in Martineau's *Illustrations*, No.25. Particularly, nouns such as 'society', 'principle', and 'class' are included. These results indicate that the central arguments based on the labour theory of value in Ricardo's *Principles* are greatly diluted in Martineau's *Illustrations*, No.25.

Next, according to the centrality analysis on the co-occurrence network in Figure 1.3, 'capital' and 'society' play the most central role in Martineau's *Illustrations*, No.25. This result reinforces the suggestion that the central arguments in Ricardo's *Principles* do not occupy the same fundamental position in Martineau's *Illustrations*, No.25.

Moreover, according to Figure 2.3 displaying the sub-graph detection of the co-occurrence network analysis, the first group of nouns (in green) in Martineau's *Illustrations*, No.25, consists of 'labourer', 'capitalist', '*population*', '*wages*', '*capital*', '*labor*', and '*proportion*'. This group includes the components (in italics) that are common to the first group of Mill's *Elements*, thus some level of correspondence with the latter is observed. In addition, the second group of nouns (in red) in Martineau's *Illustrations*, No.25, consists of '*commodity*', '*country*', '*money*', and 'place'. This group includes the components (in italics) that are common to the second group of Mill's *Elements*, again showing correspondence, to some extent, with the latter. Furthermore, the third group of nouns (in blue) in Martineau's *Illustrations*, No.25, consists of 'food', 'supply', 'demand', and 'production'. This group corresponds to the small group consisting of 'supply' and 'demand' in Mill's *Elements*. These results reflect that Martineau's *Illustrations*, No.25, has a construction of arguments that is relatively analogous to Mill's *Elements*. That being said, many small groups (in yellow, green, pink, purple, gray, and orange) are seen in Martineau's *Illustrations*, No.25, more often than in Mill's *Elements*. These results suggest a further decentralization in the construction of arguments in Martineau's *Illustrations*, No.25, more so than in Mill's *Elements*.

In this way, the results of the text-mining analysis show that, although Martineau's *Illustrations*, No.25, is to some extent, similar to Mill's *Elements*, the central concepts and the construction of the arguments differ greatly. In Martineau's *Illustrations*, No.25, the construction of the arguments with respect to the labour theory of value and the decision of corn prices --- which are crucial components in Ricardo's *Principles* --- is further diluted, and in place of these, the relation between capital and labour, or the problem of class, moves to occupy the central positions in the work.

5 Concluding remarks

5.1 Summary

In this paper, we have clarified the differences of the construction of arguments between Ricardo's *Principles*, Mill's *Elements*, and Martineau's *Illustrations*, No.25, by applying the text-mining analysis to those works. According to the results of the frequency analysis, the most frequent nouns in Ricardo's *Principles* drop down in rank in Mill's *Elements*; this pattern is advanced in Martineau's *Illustrations*, No.25, with nouns dropping down even further in rank, and in their place, other nouns occupy these high-ranking positions. Most of the nouns that see a decrease in rank correspond to the theoretical concepts regarding the decision of price and the distribution of income; most of nouns that rise in the rank correspond to realistic concepts used to express roughly a view of society. According to the co-occurrence network analysis, the nouns that play central roles in Ricardo's *Principles* are replaced by other nouns in Mill's *Elements*; those nouns are replaced yet again by other nouns in Martineau's *Illustration*, No.25. At this point, the nouns that correspond to the theoretical concepts concerning the decision of price and the distribution of income are gradually derived their roles by the nouns that correspond to the realistic concepts used to express roughly views of society. Additionally, the groups of the nouns that are connected to each other in Ricardo's *Principles* are, to some extent, maintained in Mill's *Elements*, although they exhibit slight changes; those groups are altered and greatly decentralized in Martineau's *Illustrations*, No.25. As a result, arguments constructed through strict theoretical concepts are decentralized and gradually popularized. The outcome of these considerations is seen in the process by which the specialized arguments in Ricardo's *Principles* were progressively simplified in Mill's *Elements* and decisively popularized in Martineau's *Illustrations*; in this way, Ricardian economics broadly spread out to the general public.

Therefore, the hypothesis proposed in the beginning of this paper is justified; that is, because Martineau assumed the general public as a reader of her *Illustrations*, (1) she removes technical terms and uses many popular words easily understood by the general public, and (2) she grounds her discussion in real world policies and broader worldview, rather than simply discussing basic theoretical analysis, in contrast with Ricardo and Mill. This paper demonstrates that such hypotheses were supported by the text- mining analysis.

5.2 Problems left unfinished

While the aforementioned hypotheses are supported by the text-mining analysis, they are not definitely proved solely based on this technique. Needless to say, we must examine the text data not only quantitatively, but also qualitatively, in order to sufficiently prove such hypotheses.¹¹ As previously mentioned, the text-mining analysis only provides a quantitative measure to supplement a qualitative analysis. In this sense, we have not yet fully elucidated the process of the dissemination of classical economics, and the text-mining analysis remains incomplete. Due to limits on the scope of this paper, we must leave this task to future investigations.

Despite these limitations, this study clarified certain aspects concerning the process of dissemination of classical economics, and demonstrated the suitability of text-mining analysis in studying the history of economic thought. This paper is a trial of text-mining analysis on the process of dissemination of classical economics and a preparation for further full-scale text-mining analysis.

[Acknowledgements] This work is supported by JSPS Grant-in-Aid for Scientific Research (B), Grant Number 22330064.

Notes

- 1) Such a problem is closely related to the problems studied by certain scholars as the ‘institutionalization of economics’ or the ‘history of economic policy thought’. Particularly important studies include the following: Coats (1981); Coats (1993); Furner & Supple (1990); Ikeo (2006); Nishizawa, Hattori & Kurita (1999).
- 2) The Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B), Grant Number: 22330064, Research Term: 2010-2014, Leader: Hiroyuki Shimodaira (Yamagata University), Partakers: Tatsuro Kanai (Tohoku Institute of Technology), Atsushi Komine (Ryukoku University), Shinji Fukuda (Hirosaki University), Keiko Funaki (Musashi University), Yutaka Furuya (Tohoku University), Ryo Hongo (Kansei Gakuin University), Naoki Matsuyama (University of Hyogo).
- 3) Refer to Shimodaira, Komine & Matsuyama (2012), Section 1, for the problems encountered when applying text-mining analysis to the study of the history of economic thought.
- 4) The KH Coder is a free software used for the text-mining analysis that can investigate the characteristics of the text data by extracting and aggregating words from those data. It can be downloaded online from the site (<http://khc.sourceforge.net/>). Refer to Higuchi (2012) for usage instructions.
- 5) The figure in the network of the centrality analysis is drawn based on ‘mediation centrality’, and the figure of the network of the ‘sub-graph detection’ is drawn using the ‘method based on the mediation of the co-occurrence relationship’ (Higuchi 2012: 63).
- 6) The colour of the group of words is determined by automatic processing in the KH Coder, so the colour has no inherent importance. Therefore, it is desirable that we do not rely too heavily on this classification but instead use it to assist the interpretation of the figure of the network (Higuchi 2012: 62).
- 7) We can also download the digitized text data of the literature from the website ‘The Online Library of Liberty’ (<http://oll.libertyfund.org/>). However, these data have some faults; for example, the pages of the original works are not indicated.
- 8) In the process of counting the frequency of appearance, the plural forms of the nouns are

transformed into the singular forms and are counted with the singular forms by automatic processing of the KH Coder. In addition, the British spellings are transformed into the U.S. spellings automatically; for example, ‘labour’ is transformed into ‘labor’. Because such adjustments are made automatically, we must not give the result of the analysis an excessively precise interpretation.

- 9) We can confirm the numerical value of the Jaccard coefficient of each word by using the command ‘the search for the related words’ of the KH Coder. For example, the ranking (the 1-10) of the value of the Jaccard coefficient between the highest ranking noun of frequency, ‘price’, and the other nouns in Ricardo’s *Principles* are indicated in the table below.

The ranking of the Jaccard coefficient of the noun ‘price’ in Ricardo’s *Principles*

Rank	Extracted Word	Whole	Co-occurrence	Jaccard Coefficient
1	Commodity	592	264	0.2435
2	corn	411	214	0.2246
3	profit	452	171	0.1649
4	money	369	157	0.1622
5	wages	323	136	0.1442
6	produce	366	141	0.1437
7	rise	206	119	0.1412
8	labor	514	156	0.1400
9	tax	410	133	0.1288
10	quantity	450	132	0.1229

Here, the column ‘whole’ expresses the number of the sentences in which the noun appears, and the ‘co-occurrence’ expresses the number of the sentence in which ‘price’ and the noun appear together. Based on such numerical values, the figure of the network of the ‘centricity’ and the figure of the network by the ‘sub-graph detection’ are drawn. Refer to Higuchi (2012: 49-52) for a detailed account of the command ‘the search for the related words’.

- 10) Refer to Funaki (2006), Sakurai (2003), and Uemiya (1996) for details on the influence of Martineau's Illustrations.
- 11) We can connect the quantitative analysis of the text data to the qualitative analysis by using the command ‘the KWIC concordance’ of the KH Coder. We are then able to investigate the qualitative context of the role played by the extracted words in the literature. Refer to Higuchi (2012: 44-49) for the details of that command.

References

- Coats, A.W. (ed.) 1981, *Economists in Government: An International Comparative Study*, Durham: Duke University Press.
- Coats, A.W. 1993, *The Sociology and Professionalization of Economics*, London & New York: Routledge.
- Furner, M.O. & Supple, B. (eds.) 1990, *The State and Economic Knowledge: The American and British Experiences*, Cambridge: Cambridge University Press.
- Martineau, H. 1932-34, *Illustrations of Political Economy*, 9 vols., London: Charles Fox.
- Mill, J. 1821, *Elements of Political Economy*, London: Printed for Baldwin, Cradock and Joy.
- Ricardo, D., Sraffa, P. (ed.) 1951-73, *The Works and Correspondence of David Ricardo*, 11 vols., Cambridge: Cambridge University Press.
- (Japanese)
- Funaki, K. 2006, Evaluation of Illustrations of Political Economy by Harriet Martineau in the History of Economic Thought, *Journal of Musashi University Research Center*, 16: 151-69.
- Higuchi, K. 2012, *Reference Manual for the KH Coder 2.x* (<http://khc.sourceforge.net/>).
- Ikeo, A. 2006, *Economics in Japan: The Internationalization of Economics in the Twentieth Century*, Nagoya: Nagoya University Press.
- Kida, M. 2008, *How to Use the Text Mining*, Tokyo: Hakuto Shobo.
- Jin, M. 2009, *Introduction to the Statistical Science of Text Data*, Tokyo: Iwanami Shoten.
- Muramatsu, N. & Miura, A. 2009, *Text Mining for Humanities and Social Sciences*, Tokyo: Seishin Shobo.
- Nishizawa, T., Hattori, M. & Kurita, K. (eds.) 1999, *History of Economic Policy Thought*, Tokyo: Yuhikaku.
- Shimodaira, H., Komine, A. & Matsuyama, N. 2012, An Introduction to Text Mining for the History of Economic Thought: Keynes's General Theory and its Book Reviews, *Discussion Paper Series: Research Group of Economics and Management*, No.2012-E02, Yamagata University.
- Sakurai, T. 2003, Appearance of H. Martineau's Illustration of Political Economy, *Journal of Musashi University*, 50 (2): 201-63.
- Uemiya, S. 1996, On H. Martineau's Political Economy, *Journal of Political Economy and Commercial Science*, 173 (4): 31-51.

Table 1 The highest-ranking nouns of the frequency of appearance
in Ricardo's Principles, Mill's Elements, and Martineau's Illustrations, No.25

Rank	Ricardo's Principles		Mill's Elements		Martineau's Illustrations	
	Noun	Frequency	Noun	Frequency	Noun	Frequency
1	price	1088	capital	446	labor	159
2	value	925	commodity	408	capital	148
3	commodity	750	labor	400	man	83
4	labor	684	value	341	time	80
5	capital	637	quantity	314	society	77
6	quantity	596	country	254	country	73
7	profit	573	case	242	labourer	72
8	tax	568	money	213	principle	70
9	corn	552	wages	208	production	64
10	money	500	production	194	class	60
11	country	496	produce	187	commodity	58
12	rent	472	profit	184	proportion	56
13	produce	452	tax	179	capitalist	55
14	land	432	man	170	power	54
15	wages	410	land	168	money	51
16	production	284	price	159	case	50
17	proportion	277	demand	142	demand	44
18	rise	243	exchange	130	population	44
19	demand	242	thing	127	exchange	43
20	labourer	216	portion	126	law	42
21	effect	212	labourer	122	value	42
22	landlord	202	cloth	114	food	41
23	rate	201	corn	114	supply	41
24	goods	196	metal	114	state	40
25	market	191	rent	114	increase	38
26	gold	182	goods	112	wages	38
27	trade	176	proportion	106	means	37
28	time	164	cost	98	wealth	37
29	man	160	whole	97	number	36
30	farmer	159	supply	93	people	36

Notes: The red indicates nouns included in the 10 highest-ranking in Ricardo's Principles.
The blue indicates nouns included in the 30 highest-ranking in Martineau's Illustrations only.

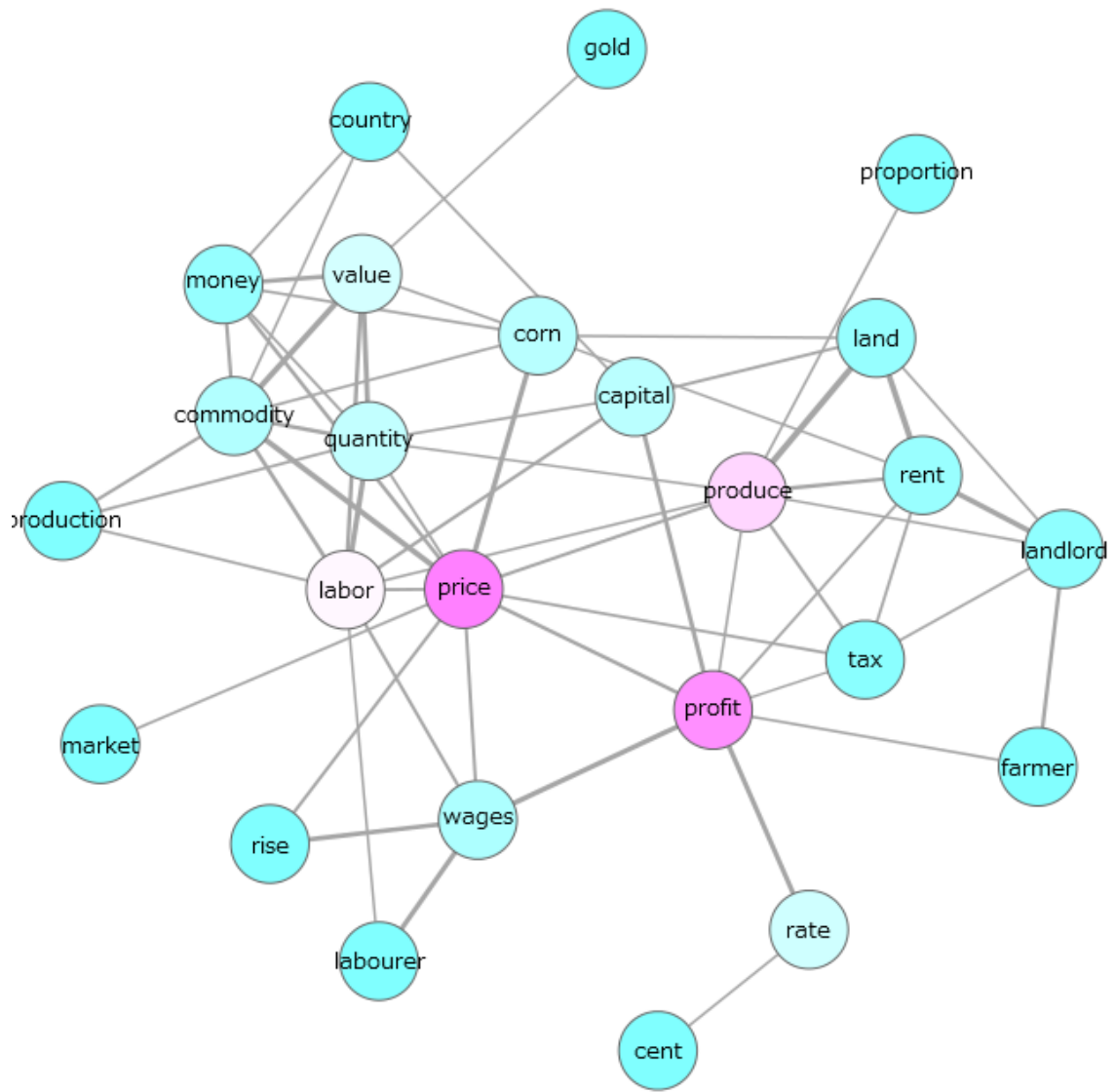


Figure 1-1

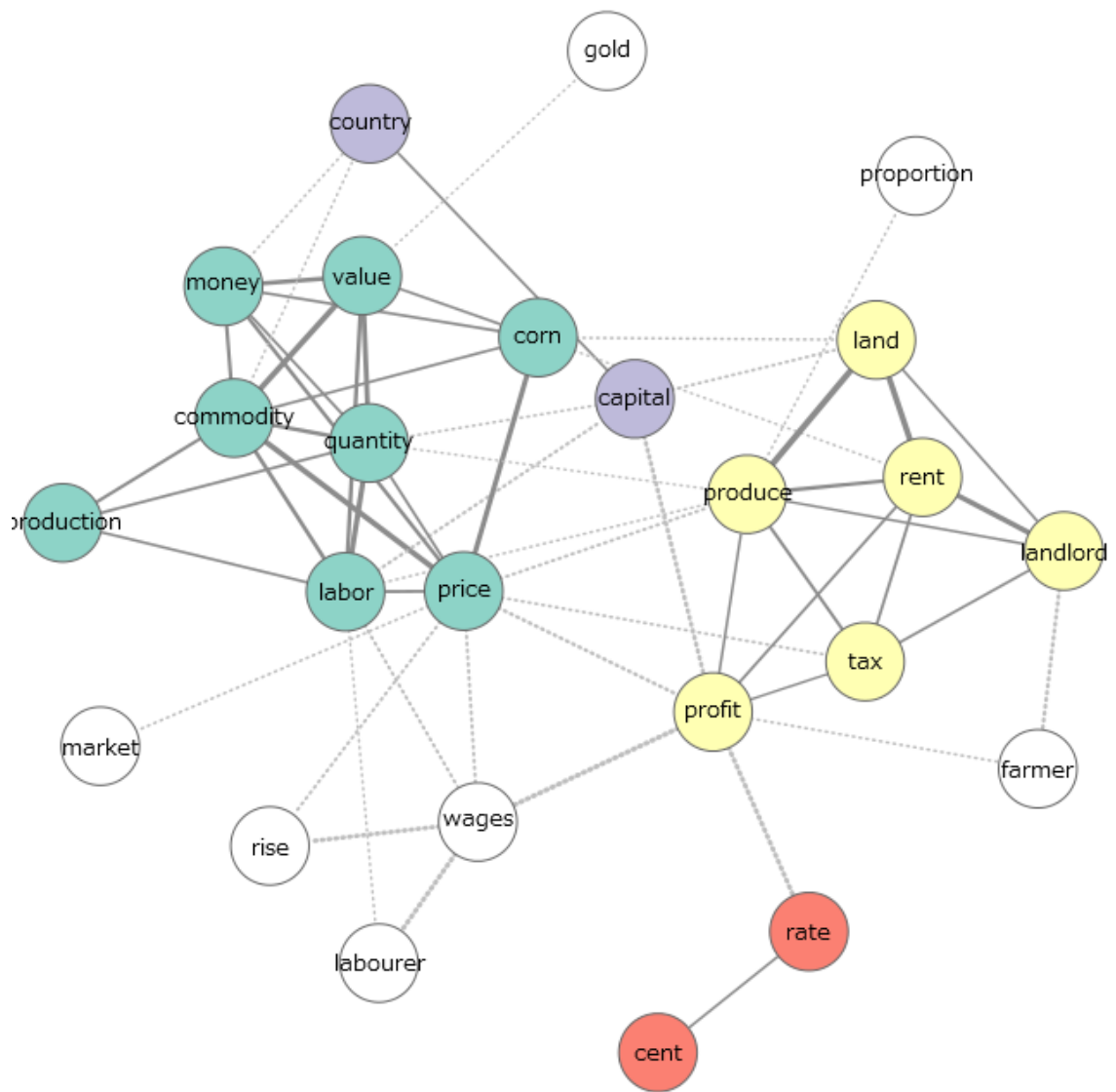


Figure 2-1

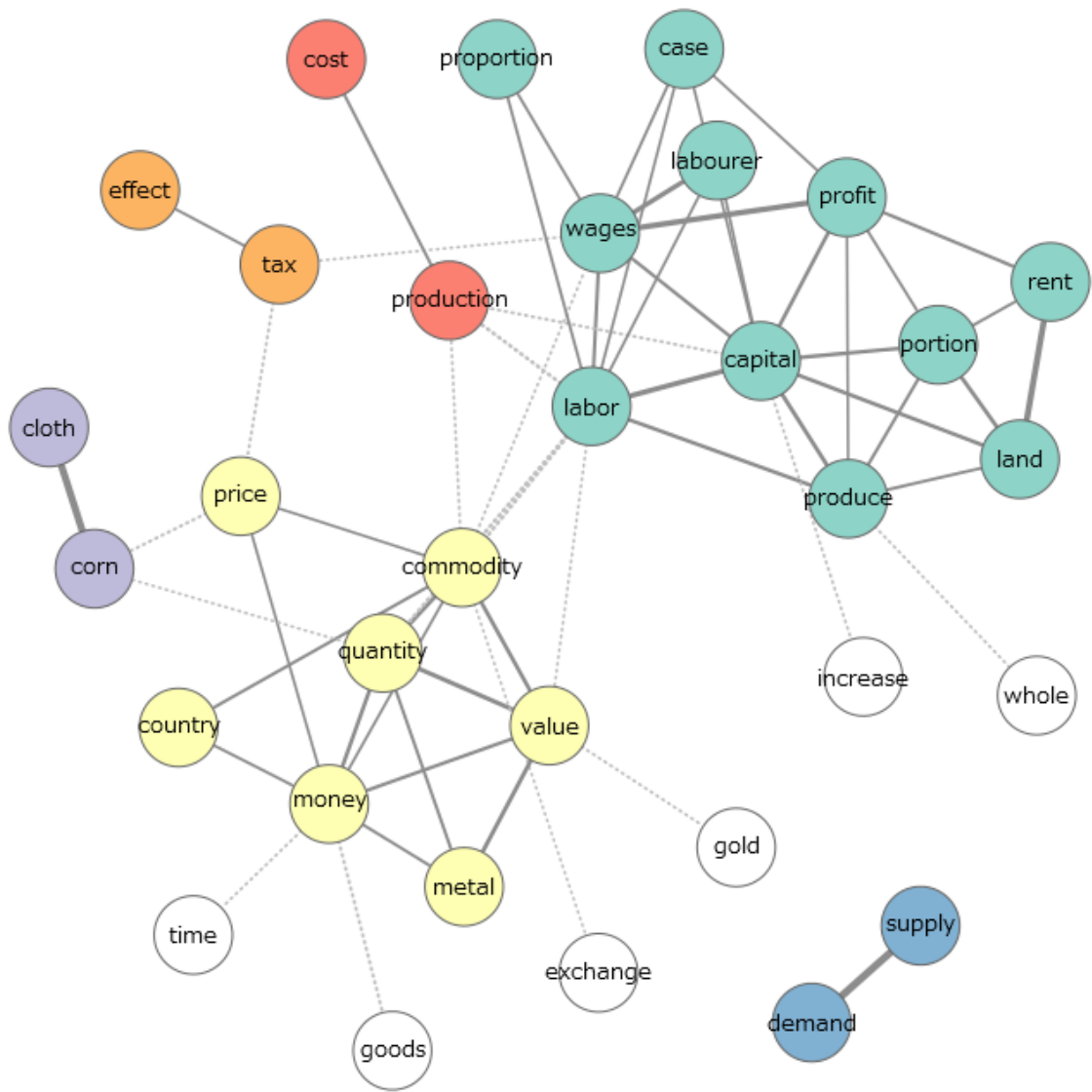


Figure 2-2



Figure 2-3