

ISSN 1881-6436

## **Discussion Paper Series**

No. 18-01

Keynesian Elements  
in Beveridge's Free Society (1944):  
A Text Mining Approach to the History of Economic Thought

**Atsushi KOMINE, Hiroyuki SHIMODAIRA**

September 2018

Faculty of Economics,  
Ryukoku University

67 Tsukamoto-cho, Fukakusa, Fushimi-ku,  
Kyoto, Japan  
612-8577

Title

**Keynesian Elements in Beveridge's Free Society (1944):  
A Text Mining Approach to the History of Economic Thought**

Authors

**Atsushi KOMINE<sup>1</sup>, Hiroyuki SHIMODAIRA<sup>2</sup>**

Abstract

This study has two conclusions. First, text mining can apply to the orthodox approach regarding HET and also be a new method to encourage heuristic knowledge; however, such application is possible only to the extent that researchers bear in mind the limitations of text mining. Second, we confirm the following two-part hypothesis by utilizing frequency, co-occurrence network, cluster, and compound word analyses, among others: (i) Beveridge's *Free Society* focuses more on insufficient effective demand, a surface cause of unemployment, than on the intricate monetary economy, which is by far an essential for Keynes [the simplification of problems], and (ii) Beveridge maintains his own views of society (which are comprehensive views from international and social perspectives) and the original (supply-side) prescription for unemployment, even after absorbing Keynesian elements [the maintenance of originality].

The two elements were in part already discussed in Komine (2007, p. 332) in approximate form. However, this study demonstrates the two elements more precisely by using numerical tables and figures, and supplementing the prior works with new findings (Beveridge's international perspective and Keynes's view of the monetary economy).

Keywords

text mining, methodology of the history of economic thought, compound analysis, full employment policy, Keynes, Beveridge

JEL classification

B22, B25, B31, B41, C18, C80

---

<sup>1</sup> Faculty of Economics, Ryukoku University (67 Fukakusa, Fushimi, Kyoto, 612-8577);  
komine@econ.ryukoku.ac.jp

<sup>2</sup> Faculty of Humanities and Social Sciences, Yamagata University (1-4-12 Kojirakawa-machi,  
Yamagata-shi, 990-8560 Japan)

6 September 2018

**Keynesian Elements**  
**in Beveridge's *Free Society* (1944):**  
**A Text Mining Approach to the History of Economic Thought**

Atsushi KOMINE (Ryukoku University)  
komine@econ.ryukoku.ac.jp  
Hiroyuki SHIMODAIRA (Yamagata University)

Version 1.9.1  
(not to be quoted)

1. Introduction
2. Text mining
  - 2.1 Definition and background
  - 2.2 The application of text mining to the history of economic thought
  - 2.3 Current studies of the digitalization technique
  - 2.4 The degree of similarity between words
3. Beveridge's *Free Society*
  - 3.1 Background in the 1930s and 1940s
  - 3.2 Frequency analysis
  - 3.3 Cluster analysis
  - 3.4 Co-occurrence network analysis
  - 3.5 Compound analysis
4. Concluding remarks

## 1. Introduction<sup>1</sup>

Our study belongs to a larger project around the question, “how did economic theories penetrate people, and then, contribute directly or indirectly to the formation of economic policies?” The project is related to research branches, such as the professionalization of economics, history of economic policy ideas, or the Keynesian Revolution in policy-making (Coats ed., 1981; Booth, 1989; and Furner & Supple eds., 1990). From 2010 onwards, our research group, based in Tohoku (northeast) Region, Japan, has attempted to clarify, qualitatively and quantitatively, the dissemination processes of British economic ideas (from 1750 to 1950), into other countries, groups, or among ordinary people.

This paper, as a case study of the project<sup>2</sup>, aims at a specific twofold target. First, in the so-called big data area, it will argue whether it is possible and suitable to apply a text mining approach to the traditional one in the HET or not, by defining a seemingly new quantitative method, namely text mining, and by pointing out its merits and demerits. Second, the paper will demonstrate how an influential theory affected contemporary leading thinkers and persons in charge of policy-making. More specifically, it will argue to what extent the elements of Keynes’s *General Theory of Employment, Interest, and Money* (*GT*, 1936) had an influence on Beveridge’s *Full Employment in a Free Society* (*FS*, 1944) in the context of accepting full employment policy by the British government.

We will validate the following two-part hypothesis by using text mining: (i) the *FS* focuses more on insufficient effective demand, a surface cause of unemployment, than on the intricate monetary economy [the simplification of problems], and (ii) Beveridge maintains his own views of society and the original prescription for unemployment, even after absorbing Keynesian elements [the maintenance of originality]. Komine (2007, p. 332) has already examined, albeit approximately and partially, the hypothesis by a traditional method.

---

<sup>1</sup> This paper is revised from the previous Japanese version, Komine & Shimodaira (2017).

<sup>2</sup> Some of our results include Shimoraida, Komine, & Matsuyama (2012); Shimodaira & Fukuda (2014); Furuya (2014); and Matsuyama (2016).

## 2. Text mining

### 2.1 Definition and background

Text mining is defined as a set of processes and methods assisted by software and hardware for a semantic and heuristic analysis of very large amounts of natural, or unstructured, text. In this regard, text mining can examine frequency, networks, patterns, trends, topics, and other characters in the data (see Wiedemann 2016, pp. 2-3; Kida, 2018, p. 8).

Until the 1990s, it was expensive to obtain information because computing power was at an insufficient level. Thus, researchers attempted to determine ‘the real world’ through as little information as possible. This was the age of the ‘classical statistical analysis’, the strong point of which was *ex-post* validation of certain phenomena. Then, from approximately the 2000s onwards, it became inexpensive to acquire even large volumes of data because computing power developed to an incredible level. Moreover, access to significant amounts of data became possible ‘thanks to ... powerful algorithms and software’ (Ignatow & Mihalcea, 2017, p. 5). Researchers no longer needed to make any special assumptions such as normal distribution or particular variance.

Text mining, which is a mixture of natural language processing and data mining in that order, has developed in the age of so-called big data. First, analysts need to engage in natural language processing (morphological and syntactic analyses). Morphological analysis means both resolution into morphemes, which are meaningful and minimum units of ‘words’, and parsing, which resolves sentences into their component parts and syntactical functions. Syntactic analysis indicates the relationship between modifiers and modificands. Second, the analysts must undertake data mining, namely frequency processing, numerous means of statistical processing (such as recurrence, principal component, correspondence, decision tree, cluster, neural networks, and latent semantic analyses), and visualization. In normal data mining, ‘data’, such as the names of commodities and numbers, must be ‘structured’ data, which consists of aggregated numeric data in tabular

form (a matrix). However, in text mining, data is necessarily text, which is natural, unstructured, and untagged, and has diverse meanings within particular contexts.

As described, text mining has dual characteristics: it stands in the middle of quantitative and qualitative analyses<sup>3</sup>. On the one hand, researchers can find out hidden patterns or networks which are automatically derived from processing big data [a data-oriented viewpoint]. On the other hand, from the patterns or networks, researchers can select significant, meaningful, and useful principles which are relevant to their own disciplines such as economics [a theory-oriented viewpoint]. At the former step, mathematical or quantitative processing helps with tools such as statistical software and algorithms can help. At the latter step, however, qualitative inference is absolutely necessary when researchers have to judge which categories are relevant, how many headlines are necessary, what words are important even if they are not frequently used, and so on. In other words, text mining is an intermediate method between an analytical focus on data (statistical processing) and one on human nature (fieldworks, case studies, and ethnography, among others). Text mining also involves an analytical focus on context by way of text.

Finally, from the very beginning of its use, text mining has had eclectic characteristics. These characteristics provide the possibility of applying text mining to the history of economic thought (HET).

## 2.2 The application of text mining to the history of economic thought

A new concept of ‘digital humanities’ now prevails in some fields; however, it has been delayed in reaching academic areas such as history, literature, the arts, and philosophy. A cause of the delay is the slow pace at which the text, which is the focal point of these areas, is digitalized. Besides, digital humanities, which naturally includes social sciences, not only utilizes digital tools and data; it also poses a challenge by eliminating the existing

---

<sup>3</sup> Regarding the relationship between the two approaches, see Brady & Collier eds. (2010) and Goertz & Mahoney (2012).

boundary lines of disciplines and thereby eventually remapping a system of new knowledge (see Burdick *et al.*, 2012, p. 1). Thus, is now the time to consider the possible applications of digital humanities, or one of its tools, text mining, for HET, a discipline which interprets large amounts of original texts from dead economists?

The application of text mining to HET could present a twofold opportunity or challenge, if we thoroughly understand its merits and demerits.

First, with regard to individual researchers, it may be possible to confirm their own conclusions or find new interpretations by applying this method. By using quantitative tools, researchers can cover wider fields of possible interpretations (namely, wider or quicker descriptions). Contemporary researchers are subject to the restrictions of their own ideologies, visions, or generally accepted ideas at that time. Automatic processing could bring their blind spots to their notice. This situation leads to heuristic knowledge. By using qualitative devices, researchers answer semantic questions based on immense and copious knowledge about cases, contexts, and history (namely, deeper or thicker descriptions). These two methods have a synergistic effect if employed well. Historians of economics are best known for their sincere interpretations of original texts. They know the importance of, for example, certain writings, documents, letters, and memos before they start text mining. Thus, they are in a good position to reconstruct chopped morphemes, which have been cut from the original text by machinery processing, into new meaningful texts.

Second, with regard to both the researchers' community and the whole world, it may be possible to deliver past original texts to current (and future) people more relevantly and objectively. Generally, researchers' processes of interpretations are apt to be in a black box if they firmly adhere to their own methodologies and ideologies. In contrast, the text-mining approach includes data set records, the clarified processes of interpretation, and numerical and visualized results. This approach, even partly at least, has a common method or viewpoint with that of theoretical or empirical economists, experimental scientists, and researchers in other fields, and general intellectual readers.

Thus, it provides a good opportunity for academicians in adjacent fields who have different methodologies to enter into the area of HET, and for historians of economics to collaborate with them in turn. Indeed, this opportunity could be one of the means to disseminate the usefulness of HET.

HET has four efficacious aspects which come from the use of text mining. First, text mining extracts ‘specific words’ within a paper or between papers as proper candidates for headlines or an index. The comparison of the extracted words drawn from automatic extraction and those from prior studies’ interpretations leads to confirmation or heuristic knowledge. Second, because stylometry is established, it is possible to identify anonymous authors. Before freedom of speech was established (for instance, during the age of mercantilism), numerous books and pamphlets appears which had no specific authors’ names attached. Third, text mining reveals differences among various editions by the same author and confirms changes in the author’s thoughts. Fourth, text mining traces certain transformations from the original text(s) into simplified and popularized discourse by policymakers and through public opinion. We have a special interest in this last aspect.

### 2.3 Current studies of the digitalization technique

Although digital humanities has become the fashion even in the social sciences, the examples relating to it are often from anthropology, education, sociology, policy studies, and business administration (Ignatow & Mihalcea, 2017, p. 3; Wiedmann, 2016, p. 1); Almost no studies consider in law and economics. This situation is a kind of mystery because, like other disciplines, law and economics regard ‘text’ as their most important data and object of study.

Despite such a trend, a few pioneering studies exist for HET. We exemplify four such studies. Wright (2016) describes relations with diverse individuals and groups beyond their own discipline in Vienna in the 1920s by using ‘social network analysis’. Claveau & Gingras (2016) describes the rise and decline of specific areas within economics from 1956 to 2014, in



combination with bibliometrics and dynamic network analysis. Binder & Jennings (2016) uses ‘topic model analysis’ to compare a currently dominant interpretation of the *Wealth of Nations*, which is deeply influenced by Cannan’s additions of headlines in his edition of the book, and the content automatically extracted from the 1784 third edition. Finally, O’Neill (2016) establishes a direct relationship between data which consists of a list of J. S. Mill’s borrowed books from the London Library and of his donated books, and his own corpus.

As far as we know, there is no study which directly applies text mining to HET. Thus, our study is among the pioneers and heralds the beginning of a process to determine with great care which aspects of the method are relevant to HET and how shortcomings may be minimized.

#### 2.4 The degree of similarity between words

Before we move on to a comparison of original texts by Keynes and Beveridge, we explain one example which indicates the transformation from original words into a quantitative index; namely, the degree of similarity between words. This index originates from the distributional hypothesis. The hypothesis states that words which are used and occur in the same contexts tend to have similar meanings (Harris, 1954, p. 157).

	have	new	drink	bottle	ride	speed	read
<b>beer</b>	36	14	72	57	3	0	1
<b>wine</b>	108	14	92	86	0	1	2
<b>train</b>	291	94	3	0	72	43	2

**Table 1** Word-context matrix

Table 1 illustrates in matrix form the number of times of neighbouring words (say, within an  $n$  words interval) that co-occur with a specific word (Okazaki, 2016, p. 190). For instance, the word ‘train’ occurs 43 times around the word ‘speed’ in a certain document. In this case, we have three ‘word vectors’ as follows:

beer (36, 14, 72, 57, 3, 0,1)  
 wine (108, 14, 92, 86, 0, 1, 2)  
 train (291, 94, 3, 0, 72, 43, 2)

Each word vector has several elements which indicate the frequency of co-occurrence near the word. This vector constitutes the ‘word-context matrix’. Further, we can mathematically define the similarity between words by utilizing the cosine of vectors. If the direction of two vectors is near, their angle becomes smaller (and the cosine of the angle becomes close to 1). Thus, we define ‘cosine similarity’ as follows (Sarkar, 2016, p. 283):

$$\text{Similarity } (U, V) = \cos \theta = \frac{U \cdot V}{|U| |V|}$$

where the numerator on the right-hand side is the inner product of the two vectors ( $U$  and  $V$ ) and the denominator is the multiplication of the size of each vector ( $\theta$  is the angle between  $U$  and  $V$ ). According to Table 1, we have two numerical values. For example:

$$\begin{aligned} & \text{Similarity (beer, wine)} \\ &= \frac{36*108+14*14+72*92+57*86+3*0+0*1+1*2}{\sqrt{36^2+14^2+72^2+57^2+3^2+0^2+1^2}\sqrt{108^2+14^2+92^2+86^2+0^2+1^2+2^2}} \\ &= \frac{15612}{\sqrt{9935}\sqrt{27725}} = 0.9406 \\ & \text{Similarity (beer, train)} = \frac{12226}{\sqrt{9935}\sqrt{100563}} = 0.386795 \end{aligned}$$

In conclusion, the similarity between ‘beer’ and ‘wine’ is 0.9406, whereas that between ‘beer’ and ‘train’ is 0.3868. It is clear that the relationship between words in certain writings can be expressed as numerical values by using a word-context matrix.

If data are quantitative or cardinal (namely, the absolute value of the data is important), cosine similarity works well. However, if data are qualitative or ordinal (namely, only the order within the data is important), the Jaccard coefficient works well (Ishida & Jin eds., 2012, p. 12):

$$\text{Jaccard coefficient} = \frac{|X \cap Y|}{|X \cup Y|}$$

(Jaccard distance = 1 – Jaccard coefficient)

where X or Y indicates the number of elements in a set, the numerator on the right-hand side is the absolute value of co-occurrence of two documents, and the denominator is the total number of the two sets.

### 3. Beveridge's *Free Society*

#### 3.1 Background in the 1930s and 1940s

Although Keynes believed in 1935 that he was ‘writing a book on economic theory which will largely revolutionise --- not ... at once but in the course of the next ten years --- the way the world thinks about economic problems’ (Keynes, 1973, vol. 13., p. 492), ‘the Treasury’s conversion to Keynesian economics was still incomplete when war broke out’ (Peden, 1979, p. 61).

In 1940, Keynes regarded the war economy from wider viewpoints, pointing out that the ‘importance of a war Budget is not because it will “finance” the war. ... Its importance is *social*: ... to do this in a way which satisfies the popular sense of social justice; whilst maintaining adequate incentives to work and economy’ (Keynes, 1979, vol. 22., p. 218; the emphasis is in the original). Keynes also agrees with Ernest Bevin, the Minister of Labour, that ‘social security must be the first object of our domestic policy after the war’<sup>4</sup>. It was natural for Keynes in March 1942 to admire Beveridge’s draft of a report on social security. Indeed, Keynes says that ‘I have read your Memoranda, which leave me in a state of wild

---

<sup>4</sup> TNA, PREM 4/100/5, “Professor Keynes’ Memorandum on War Aims”, 13 January 1941.

enthusiasm for your general scheme. I think it a vast constructive reform of real importance and am relieved to find that it is so financially possible' (Keynes, 1980, vol. 27., p. 204). Thus, Keynes, alongside James Meade and Lionel Robbins in the British Government's Economic Section, cooperated with Beveridge's creation of the report *Social Insurance and Allied Services* (published in December 1942)<sup>5</sup>.

Before finishing his report, Beveridge established his next target as the avoidance of idleness, namely the maintenance of high employment. After being banned from the government and public contact, he organized a private research group in Oxford, the members of which included Barbara Wootton, Joan Robinson, Nicholas Kaldor, and E. F. Schumacher. In the period to September 1943, Beveridge was influenced by the young scholars and absorbed Keynesian ideas from broader perspectives, saying that the 'programme ... is a programme of socialising demand rather than production. It makes possible the retention of private enterprise to discover the most efficient technical methods of production and to compete in meeting the social demand. ... The programme allows of demand being directed by social policy, putting prior human needs first'<sup>6</sup>.

The British government also published landmark white papers on budgets, social insurance, and employment in April 1941, May 1944, and September 1944 respectively.

### 3.2 Frequency analysis

The main object of our research is the original 1944 version of Beveridge's *Full Employment in a Free Society*. We concentrate on the main body of the book, excluding the notes and appendixes. The book is divided into nine parts, from the Preface and Part I, to Part VII and the Postscript. KH Coder (Ver. 2.00f), which is a free and multilingual software, recognized the main body of the work as having 3,880 sentences, 661 paragraphs, and

---

<sup>5</sup> See Komine (2018) for a full account of Beveridge as an LSE economist.

<sup>6</sup> Nuffield College Private Conferences, 1941-1955, Archive Collections, the Library of Nuffield College, Oxford. "Full Productive Employment in a Free Society", page 9, by W. H. Beveridge, 8 September 1943. Dr Uemiya (Osaka University of Economics) thankfully gave us some information about these documents.

nine parts.

The primary target of our research is the frequency of nouns. Such frequency often reflects significant ideas, concepts, and keywords. Although the software automatically counted the top 20 most frequently used words, it regarded many of them as proper nouns because of their capital letters. Before the end of World War II, quite a few words were written in capitals. Thus, we manually reclassified them into nouns if they were general words. In this regard, automatic processing is not always relevant: human eyes must always check the relevance of an entire text and its context. As Table 2 shows, the most frequent word is ‘employment’, the second ‘unemployment’, the third ‘industry’, the fourth ‘demand’, and the fifth ‘war’.

Tables 2, 3, and 4 present the differences between Keynes’s *General Theory (GT)* and Beveridge’s *Free Society (FS)*. Only five nouns, ‘employment’, ‘demand’, ‘labor’, ‘investment’, and ‘rate’ are among the 20 most frequently used in both books. There are 15 words which appear only in the *GT*, ‘interest’, ‘money’, (‘change’), ‘capital’, ‘income’, (‘cost’), (‘output’), (‘price’), (‘quality’), (‘term’), ‘level’, ‘consumption’, (‘increase’), (‘value’), and (‘efficiency’). Because the nine words in parentheses do not appear in book reviews of the *GT*<sup>7</sup>, the remaining six words are extremely important, even though the *FS* omits them. In turn, 15 nouns appear in the list of as the 20 most frequently used in both books only in the *FS*: ‘unemployment’, ‘industry’, ‘war’, ‘policy’, ‘country’, ‘outlay’, ‘man’, ‘trade’, ‘year’, ‘cent’, ‘problem’, ‘time’, ‘state’, ‘condition’, and ‘fluctuation’. These 15 words may reflect Beveridge’s age and ideals. Among them, ‘unemployment’, ‘industry’, ‘war’, and ‘policy’ have peculiar characteristics. ‘Outlay’ (which is ranked ninth), is a substitute for ‘income’ and ‘consumption’; consequently, it is also a unique term.

Noun	Raw data	ProperNoun		Noun	Correction
employment	604	50	1	employment	654
unemployment	565	2	2	unemployment	567

<sup>7</sup> See Shimodaira, Komine & Matsuyama (2012, p. 17).

industry	422	3	3	industry	425
demand	404	2	4	demand	406
war	353	26	5	labor	374
labor	348		6	war	353
policy	338	14	7	policy	352
country	295	4	8	country	299
outlay	259	1	9	outlay	260
man	237	12	10	trade	247
trade	235		11	man	237
year	222		12	year	222
investment	191	16	13	investment	207
rate	179		14	rate	179
cent	164		15	cent	164
problem	163	1	16	problem	164
time	152	2	17	time	154
state	148		18	state	148
condition	143	5	19	condition	148
fluctuation	138		20	fluctuation	138

**Table 2 The top 20 (Nouns in the FS)**

Adj	Raw data	ProperNoun		Adj	Correction
full	389	30	1	full	419
other	307		2	other	307
private	222	3	3	private	225
first	183	6	4	first	189
public	158	5	5	public	163
more	145		6	international	156
economic	141	3	7	social	151
total	140	2	8	national	150
general	137	8	9	economic	148
international	133	23	10	more	145
such	131		11	general	145

new	120	10	12	total	142
possible	116		13	such	131
unemployed	112	4	14	new	130
same	111		15	possible	116
social	101	50	16	unemployed	116
national	99	51	17	same	111
industrial	97	9	18	industrial	106
many	97		19	many	97
particular	95		20	particular	95

**Table 3 The top 20 (Adjectives in the FS)**

Table 3 presents Beveridge’s philosophy. ‘International’, ‘social’, and ‘national’ are three adjectives which are highly important, while ‘economic’ is of almost equal significance. This ranking did not take into account our manual correction of the raw data. Because the words of the headlines in the *FS*, all of which are in capitals, should be significant, it is important not to depend excessively on automatic processing.

Noun		Adjective	
rate	709	marginal	342
interest	684	other	251
employment	597	real	219
investment	526	same	192
money	457	aggregate	160
change	394	full	150
capital	392	current	148
income	385	new	146
cost	304	such	146
demand	294	equal	135
output	274	more	128
price	151	effective	126
quantity	249	economic	113
term	231	different	109

level	220	classical	102
consumption	219	general	102
increase	216	certain	94
labor	213	due	94
value	210	net	93
efficiency	202	prospective	82

**Table 4 The top 20 (Nouns and adjectives in the *GT*)**

Table 4 presents the characteristics in the *GT*. The shaded words indicate common ones in both the *GT* and *FS*. Thus, the unshaded words, such as ‘interest’, ‘money’, and ‘capital’ (noun) must be unique to Keynes. It is interesting that particular adjectives, such as ‘marginal’, ‘aggregate’, ‘effective’, and ‘classical’, imply a part of theoretical concepts.

Before this text mining analysis, as Komine (2007) says, Beveridge and Keynes shared a common concept, ‘socialisation’ [of effective demand or effective investment]. Thus, we collect data related to ‘social’. Table 5 makes it clear that related words are frequently used in the *FS*. Such use suggests that Beveridge considered economic problems from the viewpoints of society, socialization, and socialism.

Extracted	frequency	Part of speech
social	101	Adjective
SOCIAL	50	Proper Noun
society	43	Noun
socialism	10	Noun
socialization	10	Noun
socialize	4	Verb
socially	3	Adjectival Verb
socialist	3	Adjective
anti-social	1	Adjective
socialized	1	Adjective

**Table 5 Words related to ‘social’ in the *FS***



Keynes	24
Kaldor	15
Pigou	5
Beveridge	3
Bevin	2
Clay	2

**Table 6 Persons' names in the *FS***

Table 6 displays the vulnerability of text mining without employing professional knowledge about the original text. Nicholas Kaldor is famous for assisting Beveridge: he served as a member of a private investigation committee organized by the latter and wrote with his signature Appendix C, 'The Quantitative Aspects of the Full Employment Problem in Britain', in the *FS*. Thus, it is natural that eminent scholars who were hostile to Beveridge have highlighted such collaboration. For example, Hayek says:

He [Kaldor] wrote Beveridge's book on employment. ... Beveridge would have been entirely unable to write such a book.... That essay was just one that Beveridge didn't want to commit to, because he couldn't understand it. (Hayek, 1994, p. 86)

Further, Robbins states that the *FS* is proof of Beveridge's poor understanding of economics 'for it was notoriously written under advice' (Robbins, 1971, p. 136). Table 6 seems to prove their points.

However, the actual situation is more complicated. Using numerous drafts and memos from Nuffield College Private Conferences, among others, Harris (1997, pp. 432-443) explains the process of writing the *FS* in detail. According to the unpublished materials, Beveridge completely understands the essential points of the *GT*, apart from the technical aspects. Moreover, it was E. F. Schumacher, not Kaldor, who persuaded Beveridge to accept the Keynesian ideas (Harris, 1997, p. 434). Further, Beveridge's draft, 'Full

Productive Employment in a Free Society’, dated September 1943, goes beyond Keynes in a sense because he rejects the careless rhetoric of the *GT*:

... the employment secured should be productive and progressive. It excludes a solution of the problem by occupation which is merely time-wasting (digging holes and filling them) or destructive (war, armaments and Nazi drilling).

A footnote also states that in ‘the preparation of this draft for discussion frequent use has been made of a typescript Memorandum of Mr. E. F. Schumacher’<sup>8</sup>.

Because Schumacher’s name does not appear in Table 6 or even in the index of the *FS*, and Kaldor is referred to 15 times (although he appears only once in the index), Hayek and Robbins seem to be right. However, rigorous studies have revealed that Schumacher was more important during the transition of Beveridge into a Keynesian. This situation is further evidence that machinery counting and ranking may sometimes be misleading.

### 3.3 Cluster analysis

We now consider another technique, cluster analysis. As discussed in section 2.4, each item of data, or word, has a numerical value of similarity or ‘distance’. Cluster analysis is a method to divide an entire text into several clusters by grouping similar words into one cluster (Ishida & Jin eds., 2012, p. 11). Cluster analysis has two different approaches: hierarchic and non-hierarchic. With regard to the former, a small cluster is first made in order to group the nearest words. Then, a larger cluster is made to contain the smaller ones. Finally, the last cluster is established to cover the entire data. A dendrogram visualizes the hierarchic situation of each cluster as a whole. However, non-hierarchic analysis type divides the entire data into a designated number of clusters because it is very difficult to instantly analyze

---

<sup>8</sup> Nuffield College Private Conferences, 1941-1955, Archive Collections, the Library of Nuffield College, Oxford. Box 6, 6/2/92-97. “Full Productive Employment in a Free Society”, by W. H. Beveridge, 1-11, 8 September 1943.

large quantities of data comprehensively. Because a dendrogram varies

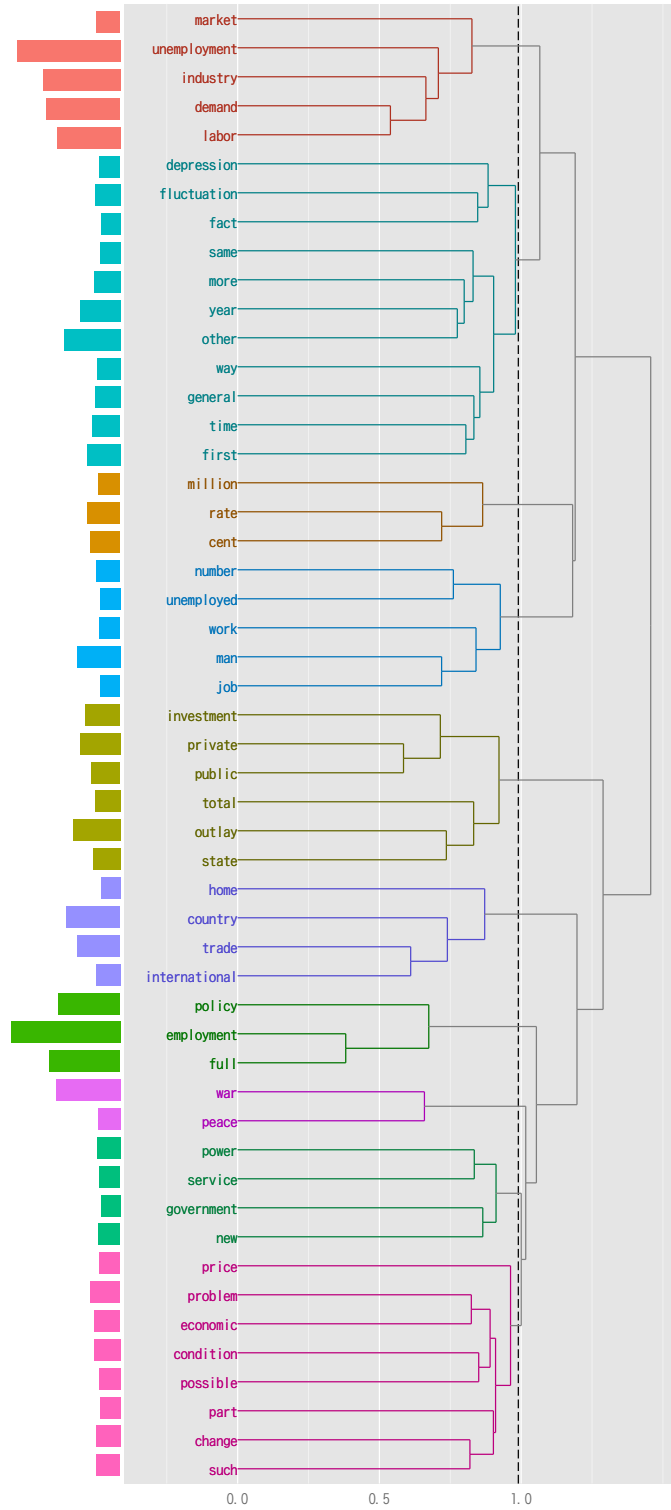


Figure 1 Hierarchic cluster analysis (10 clusters)

extensively in accordance with different definitions of ‘distance’ between data, it is safe not to take this method too seriously, and to regard it as one of hints as a supplement to other methods (Ishida & Jin eds., 2012, p. 12).

In the context of a full understanding of the aforementioned strong points and shortcomings, we apply hierarchic cluster analysis to the *FS*. The unit of grouping is ‘paragraph’, the minimum number of appearances is 105, the targets are nouns and adjectives only (excluding proper nouns), the Ward method and Jaccard coefficient are selected, and the number of clusters is ‘Auto’. The result is seven clusters. However, because this places a great many words in each cluster, 10 clusters were manually chosen instead (see Figure 1).

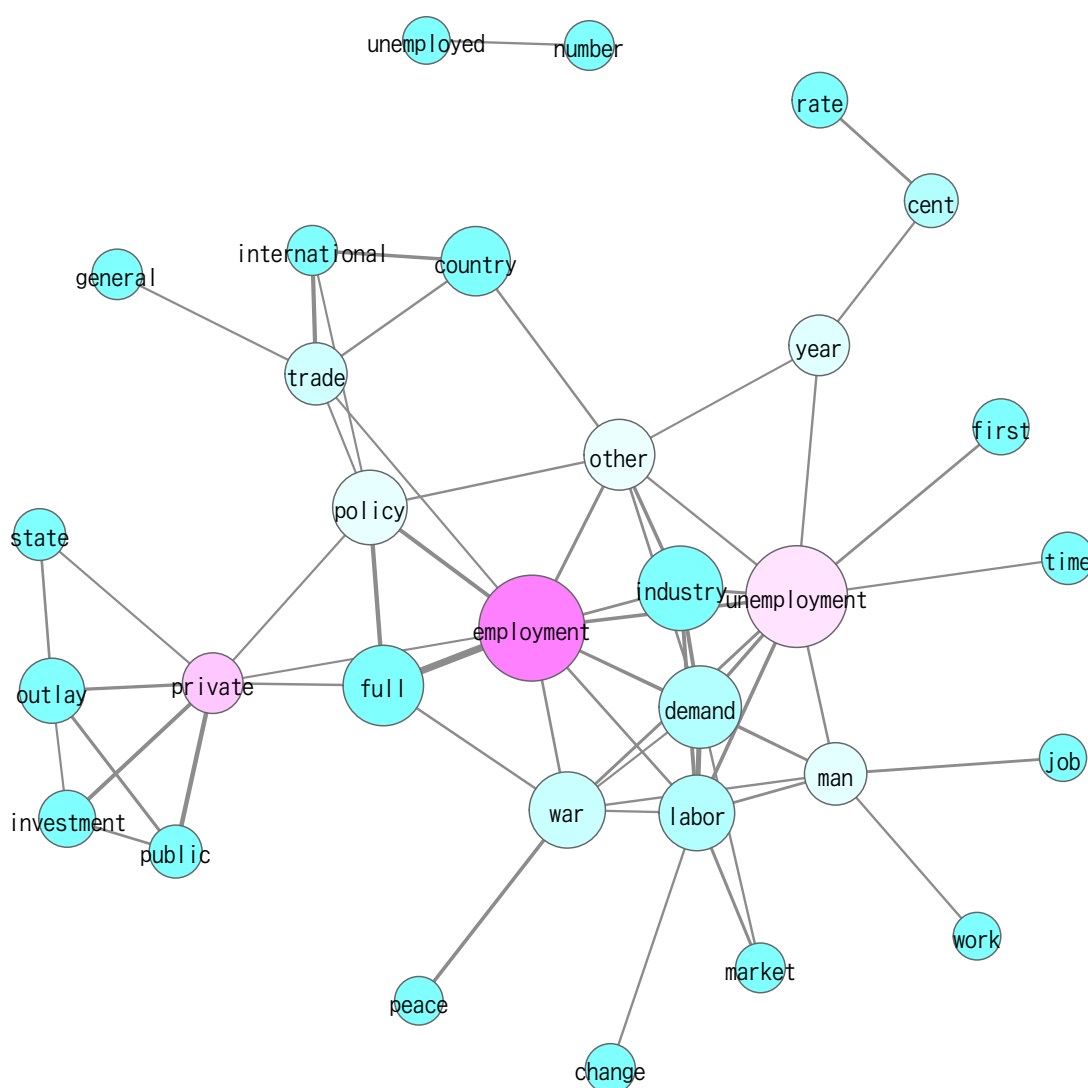
The first cluster deals with unemployment, the demand for labour, and industry; the second, depression, fluctuation, and time; the third, numerical units; the fourth, the numbers of the unemployed and jobs; the fifth, private and public investment and total outlay; the sixth, international trade; the seventh, full employment policy; the eighth, war and peace; the ninth, new services of government; the tenth, economic problems, possible changes and conditions, and prices. In conclusion, these clusters correctly summarize the main topics in the *FS*.

### 3.4 Co-occurrence network analysis

Collocation, or co-occurrence, in natural language processing refers to the phenomenon whereby a word occurs with a different word simultaneously within a given range of text. Such text is an entire document, certain paragraphs, several sentences, or  $n$  words before or after a word. With regard to the frequency of co-occurrence, it is necessary to establish a summation unit, a minimum number of appearances, and parts of speech. Because no general rule exists about how to establish these elements appropriately, researchers must proceed by trial and error.

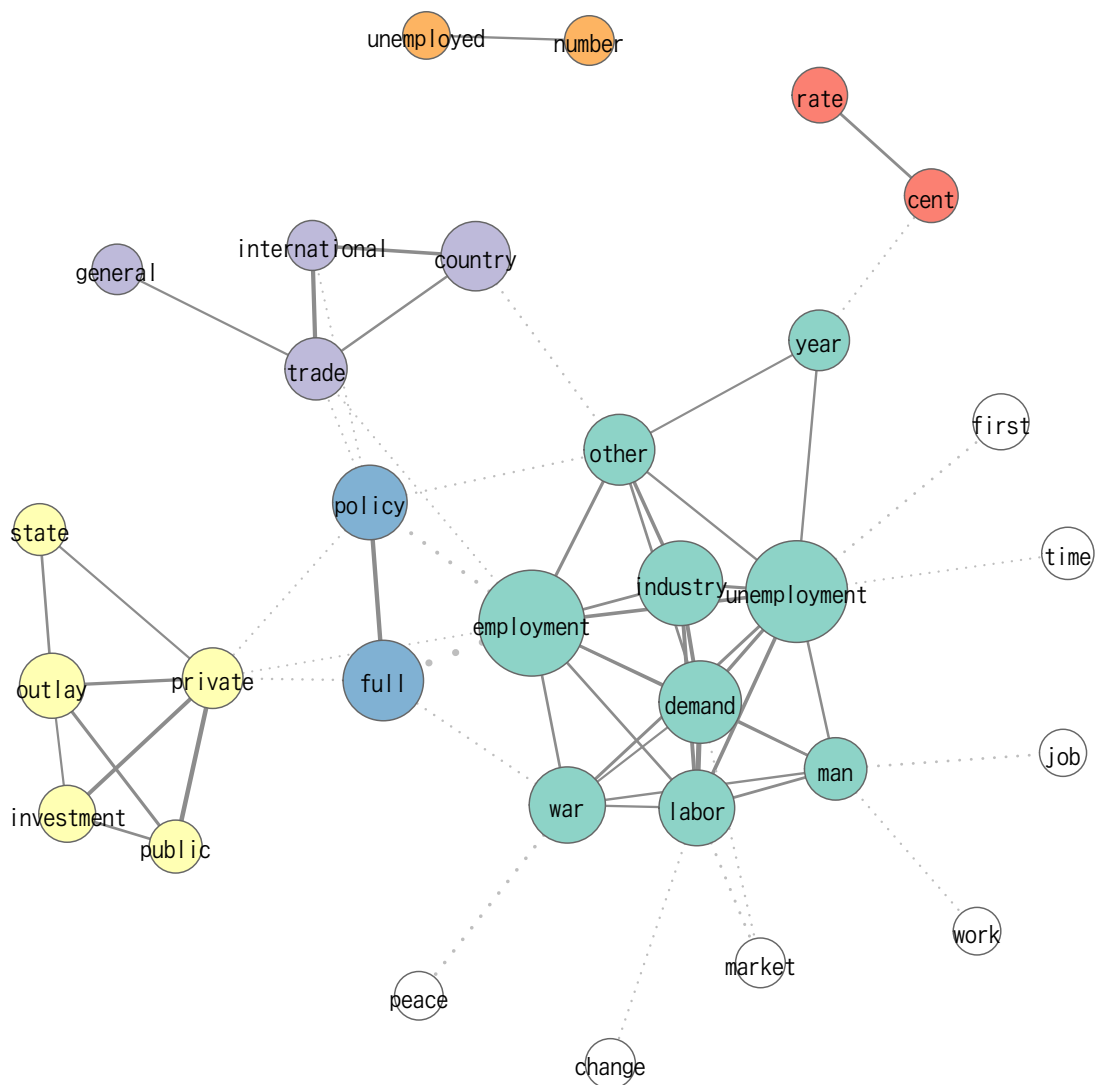
Co-occurrence network analysis is a representative method to visualize the structure of the collocation of words (Scott, 2013). It is also an

application of mathematical graph theory. In this theory, a vertex corresponds to each word, while there is an edge with the relationship to co-occurrence (or distance). Moreover, two ways can visualize the structure: the analysis of centrality (such as betweenness, degree, and eigenvector) and sub-graph detection. The former refers to the degree to which each word centrality in the network structure; the latter refers to the sort (and colour) of several words, which firmly ties one to another, in several categories (see Figures 2 and 3).



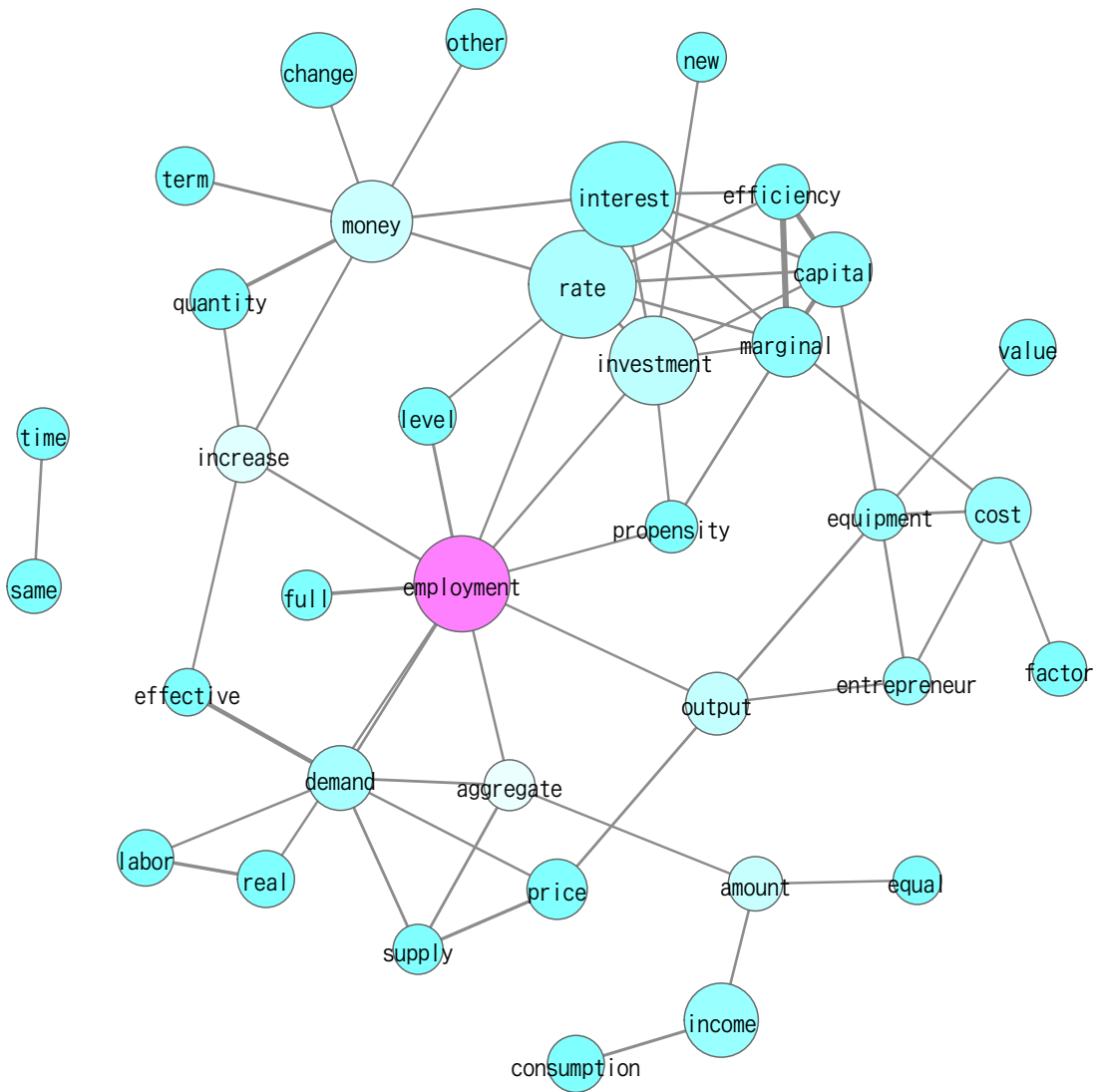
**Figure 2 Co-occurrence network analysis (centrality): the *FS***

Let us analyze the *FS* compared with the *GT*. A unit of summation is a paragraph, the minimum number of appearances is 105, and the objects of parts of speech are nouns and adjectives. In the *FS*, three words have centrality (betweenness): ‘employment’, ‘unemployment’, and ‘private’. The first and second terms are natural, whereas the third needs further explanation. Taking sub-graph detection and centrality together into consideration, the term ‘private’ might be related to not only ‘outlay’ or ‘investment’, but ‘full’, ‘employment’, ‘policy’. Beveridge seems to say that



**Figure 3 Co-occurrence network analysis (sub-graph): the *FS***

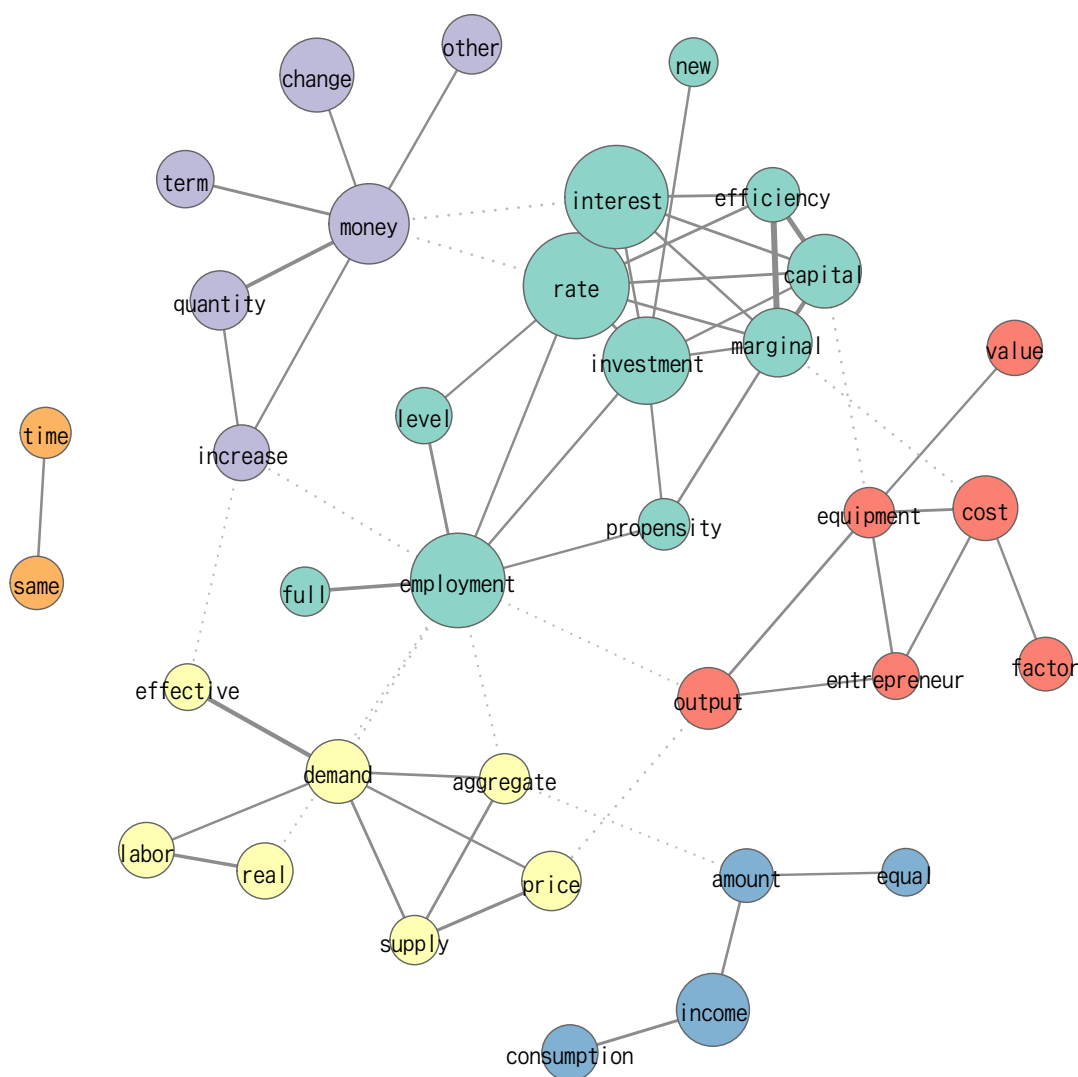
policy may have something to do with private sections of the economy. Indeed, he asserts that ‘the greater part of industry continued to be conducted by private enterprise at private risk’ (Beveridge, 1944, p. 37) and emphasizes that private companies are compatible with full employment policy (Beveridge, 1944, pp. 205-207). In this sense, the interpretation from text mining confirms traditional scholarly analysis.



**Figure 4 Co-occurrence network analysis (centrality): the *GT***

However, only one word has centrality in the *GT*. As is obvious from its title, the *GT* is concerned with the theoretical framework for determining

the level of employment. The interpretation is not inconsistent with the results of text mining. Sub-graph detection indicates the fact that there are six blocks both in the *FS* and the *GT*. However, the contents completely differ. The *FS* has the following six blocks: private and public investment, international trade, unemployment and employment, the demand for labour, the number of the unemployed, and numerical units. The *GT* has investment, marginal efficiency, and employment; the quantity of money; effective demand and supply of labour; income; cost and equipment; and time. The differences between the blocks confirm the books' characteristics. The *FS* is policy-oriented, whereas the *GT* is theory-oriented.



**Figure 5 Co-occurrence network analysis (sub-graph): the *GT***



### 3.5 Compound analysis

The next analysis is ‘TermExtract’, a system within KH Coder for extracting technical terms. A noun is apt to symbolically express a concept, an idea, or an ideal. A thinker often uses a noun not only independently but to create a new compound which includes the noun before or after the other word (Nakagawa, Yumoto, & Mori, 2003, pp. 27-28). For example, ‘full employment’ and ‘under-employment equilibrium’, which derive from ‘employment’, and ‘liquidity preference’ and ‘liquidity preference function’, which derive from ‘liquidity’. There are three steps to extract important technical terms. First, division of the text by morpheme analysis (the same process as text mining); second, counting the number of types and frequencies of words connected with a certain minimum noun; third, judging the importance of a compound created in the text (Nakagawa, Yumoto, and Mori, 2003, p. 29). Table 7 presents the 20 largest ‘scores’ (the weight of importance) of technical terms in the *FS* and *GT*.

	<i>Free Society</i>	Score	<i>General Theory</i>	Score
1	full employment	38283.435	rate of interest	59714.952
2	full employment policy	5163.317	marginal efficiency of capital	6579.732
3	international trade	3992.062	quantity of money	6536.692
4	labor market	3902.355	full employment	5807.993
5	first world war	2373.954	effective demand	4334.557
6	unemployment rate	2373.824	marginal efficiency	3995.97
7	united states	2118.38	efficiency of capital	3858.848
8	private investment	1838.44	rate of investment	3666.585
9	public outlay	1637.109	terms of money	2209.979
10	multilateral trade	1495.281	user cost	2105.101
11	unemployment insurance	1306.352	new investment	2068.328
12	mass unemployment	1276.042	real wage	1747.277
13	other countries	1233.579	real income	1619.279

14	supply of labor	1230.923	classical theory	1612.696
15	total outlay	1059.734	level of employment	1523.044
16	private outlay	1058.47	volume of employment	1511.791
17	cyclical fluctuation	1029.943	net income	1390.755
18	unemployment rates	1021.576	capital equipment	1290.096
19	social insurance	922.537	supply price	1227.051
20	total demand	873.033	aggregate investment	1184.698

**Table 7 Compound extraction (the top 20)**

This analysis produces very interesting results, which we summarize as follows. First, the top ranking compound for the *FS*, ‘full employment’, is naturally the most significant. Its score is more than seven times greater than the second-highest ranking compound. Moreover, the compounds ranked sixth, eleventh, twelfth, and eighteenth are connected with (un)employment. Second, the third highest-ranking compound, ‘international trade’, the tenth, ‘multilateral trade’, and the thirteenth, ‘other countries’ all relate to international affairs. Beveridge considers full employment policy in the light of international relations and trades. This approach contrasts with that of Keynes’s *General Theory*, which is normally regarded as a closed system. Third, as the fourth highest-ranking compound ‘labour market’ and the fourteenth, ‘supply of labour’ show, the *FS* considers supply-side problems, which Beveridge had emphasized since his earlier days in the 1900s and 1910s, and continued to do so even after absorbing Keynesian demand-side solutions.

Fourth, as the fifteenth highest-ranking compound, ‘total outlay’, the sixteenth, ‘private outlay’, and the twentieth, ‘total demand’, indicate, special compounds relate to total demand. Whether or not it is significant that these compounds are in lower positions than those relating to supply is unclear. Fifth, the fifth highest-ranking compound, ‘first world war’, the seventh, ‘united states’, the seventeenth, ‘cyclical fluctuation’, and

nineteenth, ‘social insurance’ may relate to full employment. For instance, ‘social insurance’ reflects on Beveridge’s idea that full employment is an assumption of social insurance and that the latter strengthens the former.

In addition, important noun + noun] compounds are common in Japanese, whereas adjective + noun compounds are more popular in English. The latter include ‘full employment’, ‘international trade’, ‘social insurance’, and ‘total demand’. Exceptions are ‘unemployment insurance’ for English and ‘cyclical fluctuations’ for Japanese. Thus, it is necessary to understand that the compounds of each language have their own features.

Even as a by-product, the automatic extraction of words from the *GT* is very significant. Only one compound, ‘full employment’, appears in the 20 most significant compounds in both the *FS* and *GT*. However, there are several compounds (although not the same words) about employment, investment, and total demand. The most striking feature is the order of the upper ranking: the highest-ranking compound is ‘rate of interest’, the second is ‘marginal efficiency of capital’, and the third is ‘quantity of money’. This order must reflect the causal relationship in the *GT* about determining the level of investment; namely, the rate of interest, the marginal efficiency of capital, and the quantity of money are all determinants of investment. Moreover, the tenth highest-ranking compound, ‘user cost’, the twentieth, ‘real wage’, and the fourteenth, ‘classical theory’ are all familiar to general and academic readers. What are lacking are ‘liquidity preference’ and ‘multiplier’. It is unclear why these two important compounds are missing from the 20 most significant compounds.

It is reasonable to say that the *GT* expresses theoretical concerns about monetary factors which determine the level and direction of investment. Traditionally, numerous scholars have argued about the essence of the *GT*: is it liquidity preference or multiplier (Minoguchi, 1982)? Although these two compounds have somehow not appeared from ‘TermExtract’, the 20 most significant compounds indicate the decisive importance of the ‘rate of interest’, the score of which is more than nine times that of the second-ranked ‘marginal efficiency of capital’.

‘Word Clouds’ are illustrations to visualize statistics based on frequency of appearance. Figures 6 and 7 are generated from one of the free software<sup>9</sup> of Word Clouds. It is obvious which illustration shows the *FS*.

#### 4. Concluding remarks

This study has two conclusions. These correspond to our objectives, as discussed in our introduction. First, text mining can apply to the orthodox approach regarding HET and also be a new method to encourage heuristic knowledge; however, such application is possible only to the extent that researchers bear in mind the limitations of text mining. Second, we confirm the following two-part hypothesis by utilizing frequency, co-occurrence network, cluster, and compound word analyses, among others: (i) Beveridge’s *Free Society* focuses more on insufficient effective demand, a surface cause of unemployment, than on the intricate monetary economy, which is by far an essential for Keynes [the simplification of problems], and (ii) Beveridge maintains his own views of society (which are comprehensive views from international and social perspectives) and the original (supply-side) prescription for unemployment, even after absorbing Keynesian elements [the maintenance of originality]. The two elements were in part already discussed in Komine (2007, p. 332) in approximate form. However, this study demonstrates the two elements more precisely by using numerical tables and figures, and supplementing the prior works with new findings (Beveridge’s international perspective and Keynes’s view of the monetary economy).

---

<sup>9</sup> Wordle: <http://www.wordle.net/> (access: 28 August 2017).





## Acknowledgements

This study is in part supported by the following two grants:

‘War and Peace in the History of Economic Thought: Can the Dissemination of Economics decrease International Conflicts?’, Grant-Aid for Scientific Research (B), 2016-2020; Principal Investigator: Atsushi Komine (Ryukoku University).

‘From Popularization of Economic Theory to the Creation of Economic Policy: An Empirical Study by use of Text Mining’, Grant-Aid for Scientific Research (B), 2015-2020; Principal Investigator: Hiroyuki Shimodaira (Yamagata University).

We should thank Dr Emman Quinlan, Nuffield College Library, Oxford, and Dr Robin Darweall-Smith, University College Archives, Oxford, for assisting our archival works.

## References

(The specific version for this study of text mining)

Beveridge, W. H. (2015 [1960, 1944]) *Full Employment in a Free Society: A Report*, second edition, *The Works of William H. Beveridge*, Volume 6, London and New York: Routledge. (ISBN-13: 978-1138830240)

Keynes, J. M. (1965 [1936]) *The General Theory of Employment, Interest and Money*, paperback, New York: Harcourt, Brace & World. (ISBN-13: 978-0156347112)

(General literature)

Beveridge, W. H. (1942) *Social Insurance and Allied Services*, Cmd. 6404, London: His Majesty’s Stationary Office. (一圓光彌監訳『ベヴァリッジ報告～社会保険および関連サービス』法律文化社、2014年。)

Beveridge, W. H. (1944) *Full Employment in a Free Society*, London: George Allen & Unwin Ltd. [FS]

Binder, J. M. and C. Jennings (2016) “‘A Scientifical View of the Whole’: Adam Smith, Indexing, and Technologies of Abstraction”, *Journal of English Literary History*, 83(1): 157-180; DOI: 10.1353/elh.2016.0001.

- Booth, A. (1989) *British Economic Policy, 1931-49: Was There a Keynesian Revolution?* Hemel Hempstead, Hertfordshire, U. K.: Harvester Wheatsheaf.
- Brady, H. E. and D. Collier eds. (2010) *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, second edition (first published in 2004), Maryland, U. S.: Rowman & Littlefield Publishers, Inc. (泉川泰博・宮下明聡訳『社会科学の方法論争～多様な分析道具と共通の基準』勁草書房、2008年、初版の訳)。
- Burdick, A., J. Drucker, P. Lunenfeld, T. Presner, and J. Schnapp (2012) *Digital Humanities*, Cambridge, MA: MIT Press.
- Claveau, F. and Y. Gingras (2016) “Macrodynamics of Economics: A Bibliometric History”, *History of Political Economy*, 48(4): 551-592, December 2016; DOI: 10.1215/00182702-3687259.
- Coats, A.W. ed. (1981) *Economists in Government: An International Comparative Study*, Durham, U.S.: Duke University Press.
- Furuya, Y. (2014) “An Analysis of the Dissemination of Adam Smith’s *Wealth of Nations* by use of the Text Mining”, *TERG Discussion Papers*, Faculty of Economics, Tohoku University, No. 325, November 2016. (in Japanese) (古谷豊 (2014) 「テキストマイニングを用いたスミス『国富論』普及の分析」TERG Discussion Papers (東北大学大学院経済学研究科)、No. 325、2014.11.)
- Furner, M.O. and B. Supple eds. (1990) *The State and Economic Knowledge: The American and British Experiences*, Cambridge: Cambridge University Press.
- Goertz, G. and J. Mahoney (2012) *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*, Princeton, U. S.: Princeton University Press. (西川賢・今井真士訳『社会科学のパラダイム論争～2つの文化の物語』勁草書房、2015年)。
- Hayek, F. A. (1994) *Hayek on Hayek: An Autobiographical Dialogue*, edited by S. Kresge, & L. Wenar, London: Routledge. (嶋津格訳『ハイエク、ハイエクを語る』名古屋大学出版会、2000年)。
- Harris, J. (1997) *William Beveridge: A Biography*, revised paperback edition, Oxford: Oxford University Press.
- Harris, S. (1954) “Distributional Structure”, *WORD*, 10(2-3): 146-162. DOI: 10.1080/00437956.1954.11659520.
- Ignatow, G. and R. Mihalcea (2017) *Text Mining: A Guidebook for the Social Sciences*, California, U. S.: SAGE Publications, Inc.



- Ishida, M., and M. Jin eds. (2012) *Corpus and Text Mining*, Tokyo: Kyoritsu Shuppan., Co. Ltd. (in Japanese) (石田基広・金明哲編 (2012) 『コーパスとテキストマイニング』 共立出版。)
- Kageura, K. and B. Umino (1996) “Methods of Automatic Term Recognition: A Review”, *Terminology*, 3(2): 259-289; DOI: <https://doi.org/10.1075/term.3.2.03kag>.
- Keynes, J. M. (1936) *The General Theory of Employment, Interest and Money*, London: Macmillan. [GT] (塩野谷祐一訳『雇用・利子および貨幣の一般理論』(ケインズ全集第7巻) 東洋経済新報社、1983年。)
- Keynes, J. M. (1973) *The General Theory and After: Part I, Preparation*, paperback, the Collected Writings of John Maynard Keynes, vol. XIII, London: Macmillan.
- Keynes, J. M. (1979) *Activities 1939-1945: Internal War Finance*, the Collected Writings of John Maynard Keynes, vol. XXII, London: Macmillan.
- Keynes, J. M. (1980) *Activities 1940-1946: Shaping the Post-War World: Employment and Commodities*, the Collected Writings of John Maynard Keynes, vol. XXII, London: Macmillan. (平井俊顕・立脇和夫訳『戦後世界の形成 雇用と商品～1940～46年の諸活動』(ケインズ全集第27巻) 東洋経済新報社、1996年。)
- Kida, M. (2018) *Introduction to Text Mining, New Edition: how to deal with unstructured data in the Business Management Studies*, Tokyo: Hakuto-Shobo Publishing Company. (in Japanese) (喜田昌樹『新テキストマイニング入門～経営研究での「非構造化データ」の扱い方』白桃書房、2018年。)
- Komine, A. (2007) *W. H. Beveridge in Economic Thought: A Collaboration with J. M. Keynes et al*, Kyoto: Showa-do, 2007. (in Japanese) (小峯敦『ベヴァリッジの経済思想～ケインズ等との交流』昭和堂、2007年。)
- Komine, A. (2018) “William Henry Beveridge (1879-1963)”, in R. Cord (ed.) *The Palgrave Companion to LSE Economics*, London: Palgrave Macmillan UK.
- Komine, A. and H. Shimodaira (2017) “Keynesian Elements in Beveridge’s Free Society: Quantitative and Qualitative analyses adding the text mining approach”. *Discussion Paper Series*, No. 17-01, Faculty of Economics, Ryukoku University (ISSN 1881-6436), September 2017. (in Japanese) (小峯敦・下平裕之「ベヴァリッジ『自由社会における完全雇用』のケインズの要素～テキストマイニングを加味した量的・質的分析」 *Discussion Paper Series*, No. 17-01, Faculty of Economics, Ryukoku University (ISSN 1881-6436), 2017年9月。)
- Matsuyama, N. (2016) “A Study of Text mining for Research into the History of

- Economic Thought: The case of Alfred Marshall's *Principles of Economics* (1890)", *Discussion Paper*, University of Hyogo, No.89, 2016.
- Minoguchi, T. (1982) "Some Questions about IS-LM Interpretations of *The General Theory*", *Hitotsubashi Journal of Economics*, 22(2): 16-24, 1982-02, [also in J.C. Wood ed. (1994) *John Maynard Keynes, Critical Assessment*, Second Series, vol. V, London and New York: Routledge].
- Nakagawa, H., H. Yumoto, and T. Mori (2003) "Term Extraction Based on Occurrence and Concatenation Frequency", *Journal of Natural Language Processing*, 10(1): 27-45, 2003-01-10. (in Japanese) (中川裕志・湯本紘彰・森辰則「出現頻度と連接頻度に基づく専門用語抽出」『自然言語処理』10(1): 27-45, 2003-01。)
- Okazaki, N. (2016) "Frontiers in Distributed Representations for Natural Language Processing", *Journal of the Japanese Society for Artificial Intelligence*, 31(2): 189-201. (in Japanese) (岡崎直観「言語処理における分散表現学習のフロンティア」『人工知能』(人工知能学会), 31(2): 189-201, 2016-03-01。)
- O'Neill, H. (2016) "John Stuart Mill and the London Library: A Victorian Book Legacy Revealed", *Book History*, 19: 256-283; DOI: 10.1353/bh.2016.0007
- Peden, G. C. (1979) *British Rearmament and the Treasury: 1932-1939*, Edinburgh: Scottish Academic Press.
- Robbins, L. (1971) *Autobiography of an Economist*, London: Macmillan. (田中秀夫監訳『一経済学者の自伝』ミネルヴァ書房、2009年。)
- Sarkar, D. (2016) *Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from Your Data*, California: Apress.
- Scott, J. (2013) *Social Network Analysis*, third edition, first published in 1991, London: SAGE Publications Ltd.
- Shimodaira, H., A. Komine, and N. Matsuyama (2012) "An Introduction of a text mining approach to the History of Economic Thought: Relationship between Keynes's *General Theory* and its Book Reviews", *Discussion Paper Series*, Research Group of Economics and Management, No. 2012-E02, Yamagata University. (in Japanese) (下平裕之・小峯敦・松山直樹 (2012)「経済学史研究におけるテキストマイニング分析の導入:ケインズ『一般理論』と書評の関係」Discussion Paper Series, Research Group of Economics and Management, No. 2012-E02, Yamagata University.)
- Shimodaira, H. and S. Fukuda (2014) "A Text Mining Analysis on the Dissemination

Processes of the Classical Political Economy: with regard to mainly D. Ricardo, J. S. Mill, and H. Martineau”, *Studies in the Social Sciences*, Faculty of Letters, Hirosaki University, 31: 51-66. (in Japanese) (下平裕之・福田進治 (2014) 「古典派経済学の普及過程に関するテキストマイニング分析～リカード、ミル、マーティノーを中心に」、『人文社会論叢 社会科学篇』(弘前大学人文学部) 31号: 51-66。)

Wiedemann, G. (2016) *Text Mining for Qualitative Data Analysis in the Social Sciences: A Study on Democratic Discourse in Germany*, Wiesbaden, Germany: Springer.

Wright, C. (2016) “The 1920s Viennese Intellectual Community as a Center for Ideas Exchange: A Network Analysis”, *History of Political Economy*, 48(4): 593-634; DOI: 10.1215/00182702-3687271.

(Software)

KH Coder: <http://khc.sourceforge.net/>

Wordle: <http://www.wordle.net/>

R: <http://o-server.main.jp/r/about.html>

R Commander, 1.8-x version: <http://socserv.mcmaster.ca/jfox/Misc/Rc>

TinyTextMiner βversion: <http://mtmr.jp/ttm/>

Macromill Inc.: <http://www.macromill.com/method/index.html>