

TERG

Discussion Paper No. 325

テキストマイニングを用いた
スミス『国富論』普及の分析

古谷 豊

2014年11月21日

TOHOKU ECONOMICS RESEARCH GROUP

GRADUATE SCHOOL OF ECONOMICS AND
MANAGEMENT TOHOKU UNIVERSITY
27-1 KAWAUCHI, AOBA-KU, SENDAI,
980-8576 JAPAN

1. はじめに

本稿の第一の、主たる課題はアダム・スミスの『国富論』が社会に受容され浸透していく過程でどのような変容を受けたのかについて、テキストマイニング¹を用いた分析を試みることである。第二の、副次的な課題は、この『国富論』に関する分析例を元に、経済学説史の研究にテキストマイニングという分析ツールがどの程度有用なのかという、より一般的な問いについての見通しを探ることである。

第一の課題については、本稿では頻度分析の結果を通して、『国富論』の議論が社会に普及していく過程に関して次の三点の仮説を抽出する。(1)『国富論』とその後の受容過程での言説とのあいだでは、貨幣論、とりわけ貨幣価値の議論への関心に相違があるのではないか。(2)『国富論』とその受容過程での言説とのあいだでは、三大階級のそれぞれへの関心に相違があるのではないか。(3)『国富論』とその受容過程での言説とのあいだではグレートブリテンを取り巻く国際関係への関心について相違があるのではないか。

第二の点について本稿は次のような見通しを述べる。テキストマイニングは、経済学史研究の従来の研究手法の重要性を損なうものではないものの、速やかに広く用いられるべきツールであり、今後、副次的なツールとして無視できないものとなるであろう。このように判断する理由は次の二点である。一つは経済学史研究の従来の手法では不可能だった資料の扱いができるという点。テキストマイニングは大量のテキストデータで強みを発揮するツールであり、従来の古典の読み込みという手法の限界の一つを物理的に乗り越えてくれるものである。二つはこのツールが近年登場したツールだという点。従来

¹ テキストマイニングは比較的新しい用語で定義はまだ流動的なようである。Lachenburch は 2012 年に "Text mining and analysis is a new area of research that is less than 15 years old" と書いていて (Miner 2012, xv)、Google Ngram Viewer で見ても 1997 年頃からようやく用例が増加し始めている (smoothing level 0)。Feldman (2007) の "Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, text mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns. In the case of text mining, however, the data sources are document collections, and interesting patterns are found not among formalized database records but in the unstructured textual data in the documents in these collections" (Feldman 2007, 1) の定義に見られるとおり系統だったデータの分析とは異なる自然言語のデータの分析である。英国の NaCTeM (The National Centre for Text Mining) ではテキストマイニングを三つの過程からなる分析プロセスとして、次のような定義をしている。"Text mining is the process of discovering and extracting knowledge from unstructured data. This comprises three main activities:
- Information retrieval (IR) to gather relevant texts.
- Information extraction (IE) to identify and extract entities, facts and relationships between them.
- Data mining to find associations among the pieces of information extracted from many different texts"

の研究手法を置き換えるようなものでないものの、これまでの経済学史研究の手法とは異なる、付加的な情報を与えてくれる以上は、経済学史研究のあらゆる領域にわたって一通り、どのような情報が得られるか精査する必要がある。

テキストマイニングは過度に依拠すべきものではないものの、従来の研究手法を補助するツールとして排除すべきではなく、仮説形成や検証においてツールの一つとして保持しておくべきものであろう。

2. 先行研究：著者不明原稿はスミスが書いた？

(1) 経済学史研究とテキストマイニング

経済学史研究において、下平他（2012）以前にテキストマイニングが用いられたか否かは「テキストマイニング」の定義次第であろうが、Gherity (1993)は経済学史研究にテキストマイニングを用いた先行研究として挙げるべきであろう。Gherity (1993)は匿名の文書について、著者はアダム・スミスであるとする自説を、文章の定量的な分析から補強することを試みている。著者の識別に関するこのような定量的分析については、金他（2007）が「今日のテキストマイニングの原型である」として、次のように説明している。

文体分析の研究で計量的に文章の書き手を推定・同定する本格的な研究が行われるようになったのは 19 世紀の後半からである。例えば、オハイオ州立大学の地球物理学者 Mendenhall (1887) は、単語の長さの分布に書き手の文体の特徴が現れるという研究を『サイエンス』誌に発表した。彼はディケンズ (Dickens, 1812-1870)、サッカレー (Thackeray, 1811-1863)、ミル (Mill, 1806-1873) の文章に使われた単語の長さの分布を調べ、それが作家によって異なることから、作家の文体の特徴になり得ることを示した。1960 年前後には、判別分析の手法が書き手の同定に用いられるようになった。コンピュータが自然言語を自由に扱えない時代には、必要な文体要素を目で確認しながらカウントする原始的な方法を取っていたが、その基本的な考え方は、今日のテキストマイニングの原型であると言えよう。

(金他 2007, 255。下線引用者)

Gherity (1993)は、1763 年 12 月に The Scots Magazine に掲載された著者不明原稿 Thoughts Concerning Banks, and the Paper-Currency of Scotland の著者がスミスである、と主張している。彼は三つ（ないし四つ）のアプローチでこの主張を裏付けることを試みている。第一は状況的証拠 (Circumstantial Evidence) でこの原稿に関わった銀行家たちとスミスとのつながりやスミスの行動パターンからのアプローチ、第二は文章の内容からこれはスミス本人でなければ書けなかったであろうとするアプローチ、第

三は文章のスタイルからスミスのものではないかとするアプローチである。

第三のアプローチはさらに「ソフト」と「ハード」の二つに分けられている。「ソフト」な証拠と彼が言うのは、素材の編成の仕方、装飾的な表現や頭韻の使い方、意見の表明の明瞭さなどを判断材料とするもので、これは数量など客観的に示すものではなく Gherity も言うように「多分に主観的判断が入らざるを得ないもの necessarily involves a great deal of subjective judgment」(245)で、テキストマイニングとは異なる。他方で「ハード」な証拠は文章の長さの平均値や連語の出現頻度といった数量的な計測から判断するものである。この「ハード」な証拠に関する分析が、広い意味でのテキストマイニングに含まれよう。

(2) Gherity (1993)の「ハード」な証拠分析

Gherity はスコッツ・マガジンの *Thoughts Concerning Banks* 原稿と、この原稿の異なるバージョンの原稿二種類、合計三種類の原稿についていくつかの数量的分析をして、スコッツ・マガジンの原稿のみがスミスの手になるものであって、異なるバージョンの原稿二種類をもとに、スミスが修正・再構成して書き直したのだらうと主張している。

その際に Gherity が比較対象としたのはスミスの同時期の『国富論』初期草稿と、『国富論』(刊行版)である。分析内容は文章の長さ(平均値、中央値、最大・最小値、分布)および 16 の頻出単語についてのコロケーション(共起)分析だ。これらのなかで、コロケーション分析は有意に Gherity の仮説を支持するものとはならなかったものの、文章の長さ(平均値、中央値、分布)についての分析結果は一定程度仮説に沿った結果となった。どのような分析結果だったかを見てみよう。

Gherity は文章の長さ分析については『国富論』から三カ所、『国富論』初期草稿から三カ所の合計 6 カ所(合計 400 文、15,309 語)を抽出し、スコッツ・マガジン他三種類の原稿(スコッツ・マガジンの原稿で 63 文、2,093 語)と比較している。

	平均値	中央値		平均値	中央値
『国富論』	40.8	36	スコッツ原稿	33.2	27
初期草稿	32.5	32	別バージョン 1	62.7	57
			別バージョン 2	52.3	45

結果はスコッツ・マガジンの原稿は『国富論』とその初期草稿に近い値だが他の二種類の原稿は大幅に違う、というものだった。つまり A. 三種類の原稿のなかでスコッツ・マガジンの原稿は他の二種類と書き手が異なるかもしれないこと、B. スコッツ・マガジンの原稿の書き手がスミスだったと仮定しても、文章の長さのデータに関しては特段齟齬はないこと、を読み取ることができる。もっともこれらの分析とその結果は、文章サンプルが致命的に短く、かつ文章の長さからこの原稿をスミスのものであるとすることはで

きないなど、いくつか限界を持っている。Gherity もあくまでも三つないし四つのアプローチの一つとして補足的に用いている。

3. テキストマイニングの対象・方法と結果

(1) 対象と方法

本稿は『国富論』が社会に受容され浸透されていく過程について、テキストマイニングを用いて分析を試みる。『国富論』は同時代及び後世に非常に大きな影響を及ぼした古典であるが、その過程では『国富論』の内容が偏りも変形もなくありのままに受容されていたというのはあまりに不自然な想定で、何かしらの偏りないし変形を受けつつ受容されていたのではないか。そこを検証するためには『国富論』についての書評や紹介記事など同時代人の文章を『国富論』と比較することが有効な方法の一つではないか。このような想定に基づくものである。

これは下平他（2012）のケインズ『雇用・利子及び貨幣の一般理論』と同時代の書評との比較分析と対を成す。ただし大きな違いとしては、下平他（2012）が書評類を専門的度合いで三段階に分類してこの三グループ間の比較検討をしているのに対して、本稿は書評類を分類することはせずにひとまとめにくくり、それを『国富論』のテキストと比較検討した。理由は、18世紀後半は20世紀前半とは異なり受容する側がそのように階層分化していなかったと考えられるからである。

書評類は下平他（2012）に倣い、Key Issues シリーズの Ross (1998)を用いた。外国語書評や注釈を除いた約 66,000 語が対象である。含まれる文献は次の通りである。

1. Letter from David Hume, 1 April 1776
2. Letter from Hugh Blair, 3 April 1776
3. Letter from William Robertson, 8 April 1776
4. Letter from Adam Ferguson, 18 April 1776
5. Letter from John Millar, [April 1776]
6. Letter from Thomas Pownall, 25 September 1776
7. [From review of] An Inquiry into the Nature and Causes of the Wealth of Nations (1776) William Enfield
8. [From] Observations on the Means of Exciting a Spirit of National Industry (1777) James Anderson
9. [From] Account of the Life and Writings of Adam Smith (1794/1811) Dugald Stewart
10. [From] An Inquiry into the Nature and Origin of Public Wealth (1804) James Maitland, Lord Lauderdale

11. [From] Observations on the Subjects Treated of in Dr Smith's Inquiry into the Nature and Causes of the Wealth of Nations (1817) David Buchanan
12. [From] Letter from John Macpherson 28 November 1778
13. [From] Letter to Dugald Stewart, 1795 William Petty-Fitzmaurice, Lord Shelburne
14. [From] George III's Speech, 5 December 1782
15. Speech on Preliminary Articles of Peace, 17 February 1783 William Petty-Fitzmaurice, Lord Shelburne
16. [From] Review of the Parliament of 1784 Henry Mackenzie
17. Speech on the Eden Treaty, 12 February 1787 William Pitt
18. Tribute to Smith, 17 February 1792 William Pitt
19. Letter to William Pitt, 24 October 1800 William Wyndham Grenville, Lord Grenville
20. A Short View of the Doctrine of Smith, Compared with that of the French Economists (trans. 1806/1826) Comte Germain Garnier
21. Method of Facilitating the Study of Dr Smith's Work (trans. 1806/1826) Comte Germain Garnier
22. [From] Report On Manufactures (1791) Alexander Hamilton

他方で『国富論』は全5篇のうち第1・2篇を比較対象とした。Rossが“The first two Books of WN can be regarded as Smith's greatest achievement as an analytic economist” (Ross 1998, xiii) というように、『国富論』の理論編と広く受け止められているからである²。分量は約144,000語である。

テキストマイニングのソフトウェアはKH Coder³を用いた。設定は基本的にデフォルトで、単語検出についてはStanford POS Taggerを使ったLemmatizationを選択した。また除外ワードのリストもKH Coderのリストに手を加えずに用いている。この除外ワードは、すべての単語の結果を示すとIやa、theなど分析結果として意味のないものが多く含まれてしまうためあらかじめ集計から除くものである。分析内容については、もともと媒介を挟まない直接的な分析として、頻度分析に焦点を絞ってここでは扱う。

(2) 頻度分析の一次的結果

名詞と固有名詞についての頻度分析の上位の結果をそれぞれ表1と表2に示す。二列のうち左側は『国富論』第1・2編で頻度の高い単語とその出現回数を、右側は書評類で頻度の高い単語とその出現回数を示している。

² 全5篇を比較対象としても大きな傾向は変わらない。テキストは第6版。

³ 樋口耕一が開発。樋口(2014)を参照。

表1: 名詞の出現頻度(回数)

	『国富論』1・2編		書評・書簡類			『国富論』1・2編		書評・書簡類	
1	price	878	price	303	51	thing	115	manner	53
2	labor	736	labor	264	52	person	112	work	53
3	country	624	value	189	53	paper	109	circumstance	52
4	quantity	575	trade	182	54	state	109	course	52
5	value	485	country	158	55	master	105	idea	52
6	time	437	market	157	56	kind	104	wages	52
7	silver	432	capital	145	57	consumption	103	surplus	51
8	profit	405	profit	135	58	food	101	doctrine	50
9	money	401	quantity	111	59	use	98	merchant	50
10	produce	377	commodity	108	60	sum	97	opinion	50
11	capital	372	measure	108	61	case	96	consumption	49
12	land	342	man	107	62	landlord	96	government	49
13	year	339	time	106	63	effect	95	rent	49
14	stock	317	year	105	64	hand	95	circulation	48
15	trade	314	money	104	65	cattle	94	degree	48
16	market	289	state	104	66	material	94	branch	47
17	proportion	280	wealth	103	67	town	92	importation	47
18	wages	272	bounty	95	68	merchant	91	stock	47
19	gold	267	thing	94	69	cent	86	view	47
20	people	243	article	92	70	law	86	improvement	46
21	rent	234	produce	91	71	circulation	85	reasoning	46
22	corn	226	commerce	89	72	business	83	tax	46
23	commodity	209	effect	89	73	circumstance	77	scarcity	45
24	man	206	case	88	74	exchange	76	way	45
25	demand	193	corn	87	75	subsistence	76	deposit	43
26	expense	193	industry	87	76	farmer	75	reason	42
27	goods	177	nature	87	77	operation	73	supply	41
28	shilling	175	advantage	85	78	ounce	73	mankind	40
29	number	171	principle	85	79	advantage	70	means	40
30	employment	170	monopoly	76	80	cultivation	70	order	40
31	bank	162	operation	74	81	rise	70	business	39
32	improvement	162	object	70	82	consequence	68	law	39
33	industry	162	book	69	83	wheat	68	exportation	38
34	revenue	157	community	68	84	example	67	author	37
35	metal	154	proportion	67	85	nation	67	paper	37
36	order	154	use	66	86	wealth	67	plenty	37
37	society	151	farmer	62	87	course	66	revenue	37
38	place	149	place	62	88	day	66	colony	36
39	rate	147	nation	60	89	cause	65	truth	36
40	coin	146	society	60	90	deal	65	cause	35
41	pound	145	power	59	91	weight	65	exchange	35
42	work	143	increase	58	92	period	64	expencc	35
43	manner	141	land	58	93	way	64	employment	34
44	sort	135	rate	56	94	competition	63	end	34
45	workman	133	subject	56	95	maintenance	63	progress	34
46	occasion	130	consequence	54	96	nature	62	chap	33
47	account	125	demand	54	97	provision	62	rise	33
48	increase	123	fact	54	98	agriculture	61	source	33
49	labourer	122	home	54	99	home	58	class	32
50	century	115	grain	53	100	note	58	knowledge	31

表2: 固有名詞の出現頻度(回数)

『国富論』1・2編		書評・書簡類		『国富論』1・2編		書評・書簡類			
1	Europe	148	Smith	70	21	Henry	16	United	9
2	England	136	Britain	52	22	Indies	16	WILLIAM	9
3	Scotland	120	Mr	52	23	Edward	15	Commerce	7
4	London	78	Great	42	24	Portugal	15	ADAM	6
5	France	57	Dr	32	25	Holland	14	Germany	6
6	America	56	England	29	26	II	13	Labor	6
7	Britain	51	France	29	27	Elizabeth	12	West	6
8	Great	50	Europe	27	28	Poland	12	D	5
9	English	39	America	21	29	Ireland	11	Edinburgh	5
10	Mr	38	Sir	17	30	Rent	11	James	5
11	China	35	Nations	14	31	Tower	11	London	5
12	Scotch	28	B	12	32	Gold	10	M	5
13	Edinburgh	27	Hume	11	33	Indostan	10	Scotland	5
14	India	25	Inquiry	11	34	Paris	10	Treatise	5
15	Peru	25	Lord	11	35	B	9	Bacon	4
16	III	24	Necker	11	36	Bengal	9	Farmer	4
17	Spain	24	English	10	37	Romans	9	Indies	4
18	North	20	WEALTH	10	38	St	9	Ireland	4
19	East	19	Public	9	39	Virginia	9	Joseph	4
20	Bank	16	States	9	40	Fleetwood	8	RENT	4

ここからまず三つの点が指摘できる。第一に、名詞の出現頻度が高い単語は、『国富論』と書評・書簡類とで類似性が極めて高いことである。出現頻度上位5つの単語のうち4つが共通しており、上位2つの単語は順位も一致している。出現頻度の順位が類似しているだけではない。それらの単語の出現頻度の比率も、極めて類似している。最も頻度の高い price の出現頻度は『国富論』で 878 回、書評・書簡類で 303 回なので、その比率は 878 回/303 回=2.9 である。2 番目の labor⁴ の場合は 736 回/264 回=2.8 だ。さらに上位5つの単語のうち共通する4つの単語の合計出現頻度で見ても、2723 回/914 回=3.0。2.9 と 2.8、3.0 と、いずれも近似している。このように出現頻度の高い単語の順位と、頻度の比率とがともに極めて近似しているという事実は、書評・書簡類の段階では『国富論』を比較的歪み無く伝えているという可能性を示唆している。

第二に固有名詞の出現頻度では、単語が国名と人名に集中するため、『国富論』と書評・書簡類との類似性が(普通)名詞の場合よりさらに高くなるかと思いきや、必ずしもそうとはいえないこと。

第三に、出現頻度が低くなるにつれてその差も小さくなり、あまり有意には比較できな

⁴ 『国富論』や書評・書簡類では labour。KH Coder と Stanford POS Tagger で同じ単語として処理している模様。

いこと。

そこで次に出現頻度の高い単語を中心に、『国富論』と書評・書簡類との両者をより詳しく比較して、その相違点を浮かび上がらせてみよう。

(3) 頻度分析結果を比較する

『国富論』と書評・書簡類とで、名詞の出現頻度上位の単語が共通している一方で、個別の単語を見ていくと両者でかなり異なるものもある。また固有名詞では『国富論』と書評・書簡類とで出現頻度の類似性がより低いように見える。これらの相違点を浮かび上がらせるために表 1 と表 2 をそれぞれ表 3-a, b と表 4-a, b に加工した。

ここでは、それぞれの単語の出現頻度（回数）そのものではなくて、それぞれの単語の出現頻度が『国富論』と書評・書簡類とでどう違うかということ、出現頻度の比率で示そうとした。その際、前節で見たように、出現頻度の高い単語の順位と比率が極めて近似しているという点に着目して便宜的に次のような指標を置いた。Price と labor はそれぞれ『国富論』と書評・書簡類とで出現頻度 1 位、2 位で、その比率も 2.9 と 2.8 と近似している。これら二つの単語を合わせた比率、 $(878 \text{ 回} + 736 \text{ 回}) / (303 \text{ 回} + 264 \text{ 回}) = 2.84656$ をここでのいわば「標準的な比率」ということにする。そして各単語の出現頻度が『国富論』と書評・書簡類とでどれくらい多いのか・少ないのかを、このものさしで測るのである。

例えば price は『国富論』で 878 回、書評・書簡類で 303 回なのでその比率は $878/303 = 2.89769$ で、「標準的な比率」より若干高い。この 2.89769 を「標準的な比率」で割ると、1 より高い値である 1.02 となる。他方で value という単語は『国富論』で 485 回、書評・書簡類で 189 回なので比率は 2.566、「標準的な比率」で割ると、1 より低い値である 0.9 となる。つまり『国富論』と書評・書簡類での出現傾向を比べた場合、value は price や labor と比較して書評・書簡類でよく登場する名詞であるということになる。

表 3-a は『国富論』で頻出上位 100 の名詞のそれぞれが、書評・書簡類での出現頻度と比べてどれくらい頻出かを示す。第一列に単語、第二列に『国富論』での出現頻度、第三列に書評・書簡類での出現頻度、第四列に「標準的な比率」と比べた比率を挙げる。

表 3-b は同じ事を書評・書簡類の側から見ている。書評・書簡類で頻出上位 100 の名詞のそれぞれが、『国富論』での出現頻度と比較してどの程度頻出なのかを示す。第四列の比率が高いということはその単語が相対的に書評・書簡類でよく登場するということの意味する。

表 4-a, b は同じ事を固有名詞の頻出上位 40 で示している。

表 3-a 『国富論』で頻度上位 100 の名詞の比較

	WN	書評		
price	878	303	1.02	
labor	736	264	0.98	
country	624	158	1.39	
quantity	575	111	1.82	
value	485	189	0.90	
time	437	106	1.45	
silver	432	13	11.67	
profit	405	135	1.05	
money	401	104	1.35	
produce	377	91	1.46	
capital	372	145	0.90	
land	342	58	2.07	
year	339	105	1.13	
stock	317	47	2.37	
trade	314	182	0.61	
market	289	157	0.65	
proportion	280	67	1.47	
wages	272	52	1.84	
gold	267	9	10.42	
people	243	28	3.05	
rent	234	49	1.68	
corn	226	87	0.91	
commodity	209	108	0.68	
man	206	107	0.68	
demand	193	54	1.26	
expense	193	12	5.65	
goods	177	17	3.66	
shilling	175	12	5.12	
number	171	18	3.34	
employment	170	34	1.76	
bank	162	7	8.13	
improvement	162	46	1.24	
industry	162	87	0.65	
revenue	157	37	1.49	
metal	154	9	6.01	
order	154	40	1.35	
society	151	60	0.88	
place	149	62	0.84	
rate	147	56	0.92	
coin	146	8	6.41	
pound	145	3	16.98	
work	143	53	0.95	
manner	141	53	0.93	
sort	135	11	4.31	
workman	133	5	9.34	
occasion	130	27	1.69	
account	125	21	2.09	
increase	123	58	0.75	
labourer	122	23	1.86	
century	115	10	4.04	
thing	115	94	0.43	
person	112	17	2.31	
paper	109	37	1.03	
state	109	104	0.37	
master	105	2	18.44	
kind	104	24	1.52	
consumption	103	49	0.74	
food	101	11	3.23	
use	98	66	0.52	
sum	97	10	3.41	
case	96	88	0.38	
landlord	96	5	6.74	
effect	95	89	0.37	
hand	95	30	1.11	
cattle	94	12	2.75	
material	94	28	1.18	
town	92	2	16.16	
merchant	91	50	0.64	
cent	86	7	4.32	
law	86	39	0.77	
circulation	85	48	0.62	
business	83	39	0.75	
circumstance	77	52	0.52	
exchange	76	35	0.76	
subsistence	76	13	2.05	
farmer	75	62	0.42	
operation	73	74	0.35	
ounce	73	0	0.00	
advantage	70	85	0.29	
cultivation	70	11	2.24	
rise	70	33	0.75	
consequence	68	54	0.44	
wheat	68	11	2.17	
example	67	9	2.62	
nation	67	60	0.39	
wealth	67	103	0.23	
course	66	52	0.45	
day	66	11	2.11	
cause	65	35	0.65	
deal	65	4	5.71	
weight	65	5	4.57	
period	64	23	0.98	
way	64	45	0.50	
competition	63	29	0.76	
maintenance	63	4	5.53	
nature	62	87	0.25	
provision	62	18	1.21	
agriculture	61	22	0.97	
home	58	54	0.38	
note	58	3	6.79	

表 3-b 書評・書簡類で頻度上位 100 の名詞の比較

	書評	WN			書評	WN		
price	303	878	0.98		manner	53	141	1.07
labor	264	736	1.02		work	53	143	1.06
value	189	485	1.11		circumstance	52	77	1.92
trade	182	314	1.65		course	52	66	2.24
country	158	624	0.72		idea	52	3	49.34
market	157	289	1.55		wages	52	272	0.54
capital	145	372	1.11		surplus	51	50	2.90
profit	135	405	0.95		doctrine	50	1	142.33
quantity	111	575	0.55		merchant	50	91	1.56
commodity	108	209	1.47		opinion	50	22	6.47
measure	108	47	6.54		consumption	49	103	1.35
man	107	206	1.48		government	49	38	3.67
time	106	437	0.69		rent	49	234	0.60
year	105	339	0.88		circulation	48	85	1.61
money	104	401	0.74		degree	48	52	2.63
state	104	109	2.72		branch	47	51	2.62
wealth	103	67	4.38		importation	47	31	4.32
bounty	95	21	12.88		stock	47	317	0.42
thing	94	115	2.33		view	47	8	16.72
article	92	17	15.40		improvement	46	162	0.81
produce	91	377	0.69		reasoning	46	1	130.94
commerce	89	57	4.44		tax	46	47	2.79
effect	89	95	2.67		scarcity	45	51	2.51
case	88	96	2.61		way	45	64	2.00
corn	87	226	1.10		deposit	43	0	*∞
industry	87	162	1.53		reason	42	39	3.07
nature	87	62	3.99		supply	41	20	5.84
advantage	85	70	3.46		mankind	40	4	28.47
principle	85	23	10.52		means	40	53	2.15
monopoly	76	14	15.45		order	40	154	0.74
operation	74	73	2.89		business	39	83	1.34
object	70	25	7.97		law	39	86	1.29
book	69	28	7.01		exportation	38	36	3.00
community	68	0	*∞		author	37	12	8.78
proportion	67	280	0.68		paper	37	109	0.97
use	66	98	1.92		plenty	37	23	4.58
farmer	62	75	2.35		revenue	37	157	0.67
place	62	149	1.18		colony	36	43	2.38
nation	60	67	2.55		truth	36	2	51.24
society	60	151	1.13		cause	35	65	1.53
power	59	57	2.95		exchange	35	76	1.31
increase	58	123	1.34		expenditure	35	0	*∞
land	58	342	0.48		employment	34	170	0.57
rate	56	147	1.08		end	34	35	2.77
subject	56	19	8.39		progress	34	40	2.42
consequence	54	68	2.26		chap	33	0	*∞
demand	54	193	0.80		rise	33	70	1.34
fact	54	10	15.37		source	33	12	7.83
home	54	58	2.65		class	32	28	3.25
grain	53	20	7.54		knowledge	31	16	5.52

表 4-a 『国富論』で頻出上位 40 の固有名詞の比較

	WN	書評	
Europe	148	27	1.93
England	136	29	1.65
Scotland	120	5	8.43
London	78	5	5.48
France	57	29	0.69
America	56	21	0.94
Britain	51	52	0.34
Great	50	42	0.42
English	39	10	1.37
Mr	38	52	0.26
China	35	0	* ∞
Scotch	28	2	4.92
Edinburgh	27	5	1.90
India	25	3	2.93
Peru	25	0	* ∞
III	24	2	4.22
Spain	24	4	2.11
North	20	3	2.34
East	19	1	6.67
Bank	16	0	* ∞
Henry	16	1	5.62
Indies	16	4	1.41
Edward	15	1	5.27
Portugal	15	2	2.63
Holland	14	1	4.92
II	13	1	4.57
Elizabeth	12	0	* ∞
Poland	12	0	* ∞
Ireland	11	4	0.97
Rent	11	4	0.97
Tower	11	0	* ∞
Gold	10	1	3.51
Indostan	10	0	* ∞
Paris	10	1	3.51
B	9	12	0.26
Bengal	9	0	* ∞
Romans	9	0	* ∞
St	9	0	* ∞
Virginia	9	4	0.79
Fleetwood	8	0	* ∞

表 4-b 書評・書簡類で頻度上位 40 の固有名詞の比較

	書評	WN	
Smith	70	0	* ∞
Britain	52	51	2.90
Mr	52	38	3.90
Great	42	50	2.39
Dr	32	3	30.36
England	29	136	0.61
France	29	57	1.45
Europe	27	148	0.52
America	21	56	1.07
Sir	17	0	* ∞
Nations	14	1	39.85
B	12	9	3.80
Hume	11	3	10.44
Inquiry	11	3	10.44
Lord	11	0	* ∞
Necker	11	0	* ∞
English	10	39	0.73
WEALTH	10	0	* ∞
Public	9	0	* ∞
States	9	0	* ∞
United	9	0	* ∞
WILLIAM	9	8	3.20
Commerce	7	0	* ∞
ADAM	6	0	* ∞
Germany	6	1	17.08
Labor	6	6	2.85
West	6	8	2.13
D	5	1	14.23
Edinburgh	5	27	0.53
James	5	4	3.56
London	5	78	0.18
M	5	1	14.23
Scotland	5	120	0.12
Treatise	5	0	* ∞
Bacon	4	0	* ∞
Farmer	4	0	* ∞
Indies	4	16	0.71
Ireland	4	11	1.04
Joseph	4	0	* ∞
RENT	4	11	1.04

このように『国富論』と書評・書簡類との出現頻度の比率を比較すると、さらにいくつかの傾向が浮かび上がってくる。第一に、表 3-b で比率のとくに高い名詞を見ていくと、wealth, article, nature, principle, object, book, subject, fact, idea, doctrine, opinion, view, reasoning, author, truth, source などの論評に伴う単語や書名に含まれる単語が多い。これはとても自然なことであるとともに、『国富論』の受容過程での変容を表すものとはいえない。

第二に、表 3-a で比率の高い名詞を見ると、議論そのものに関わるような単語が多く見られる。こと回数に関する限りでは、これらの単語に関わる議論が『国富論』でなされているにもかかわらず、書評・書簡類ではそれに見合った取り上げ方はされていないといえることができる。これらについてはさらに検証が求められる。

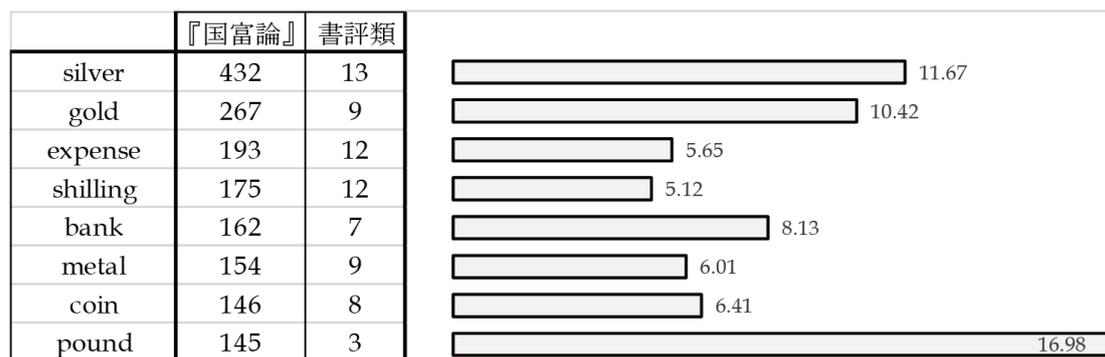
第三に、固有名詞では（普通）名詞とは異なって、『国富論』と書評・書簡類とで片方には使われているのにもう一方には一回も使われてないという単語が多くある。

そして今回の結果が興味深いのは、『国富論』の議論を知っている我々の立場から検証すると、主観的な判断を差し挟まずとも、かなり明瞭に、議論の内容と噛み合った形での相違点が出ているという点である。これらを以下に示していこう。

4. 頻度分析から得られる三つの仮説

(1) 貨幣論への関心の相違

表 3-a で比率の高いものを見ていこう。「標準的な比率」の 5 倍以上のものを順に挙げていくと silver, gold, expense, shilling, bank, metal, coin, pound, ... となる。Pound は貨幣の単位だけでなく重量の単位の pound も含まれているが、expense を除くといずれもスミスの貨幣論に関わる名詞である。



この事実から、「『国富論』が社会に受容されていく過程で、貨幣・金融についての議論が、スミスが意図していた重要性ほどには伝えられていない」という仮説を導くことができる。いうまでもなくテキストマイニングが示す事実と、この仮説との間にはまだ多くの距離がある。これらは伝統的な研究手法も含めて別途、多角的に検証しなければ埋めることはできない。ただしテキストマイニングで得られた事実がスミスの議論のくくりとかなり重なる顕著な傾向を指し示していて、そこから有意な仮説を導出することが

できたということは、テキストマイニングが分析ツールとして一定の有用性を持つことを示しているといえよう。

(2) 資本家・地主・労働者への関心の相違

スミスは『国富論』第1編第8～11章で労働者・資本家・地主へ賃金・利潤・地代がどのように配分されるか議論している。表3-aを見ると、wages, profit, rentとそれぞれの受け取り主体である labourer, farmer, landlord の出現頻度の平仄がかなり合っていることが分かる。

	『国富論』	書評類
wages	272	52
profit	405	135
rent	234	49

Category	Ratio
wages	1.84
profit	1.05
rent	1.68

	『国富論』	書評類
labourer	122	23
farmer	75	62
landlord	96	5

Category	Ratio
labourer	1.86
farmer	0.42
landlord	6.74

つまり『国富論』と比べて書評・書簡類は、利潤とその受け取り主体については回数多く言及するものの、賃金・地代とそれぞれの受け取り主体についてはあまり言及していない、というようにまとめられる。この事実から、「『国富論』が社会に受容されていく過程で、三大階級とその収入のうち、資本家とその利潤についての関心が高くそれに偏して受容されていった」という仮説を導くことができる。もちろん、これもテキストマイニングの事実から実証されたというのではなく、テキストマイニングの事実から導かれる仮説である、というに過ぎない。

(3) 国際関係への関心の相違

固有名詞については、国名・地名に着目して表4-a, bを見ると、言及の比率にかなりの相違があることが分かる。「標準的な比率」を一つの尺度として、二つのグループに分類して見よう。

『国富論』の方が相対的に言及回数の多いもの：

Europe, England, Scotland, London, China, Edinburgh, India, Peru, Spain, Indies, Portugal, Holland, Poland, Indostan, Paris, Bengal, Virginia

書評・書簡類の方が相対的に言及回数の多いもの：

France, America, Britain, Ireland, Virginia, United States, Germany, ただし America, Ireland はほぼ同等

ここからは(1)や(2)ほどには自然な仮説を導くことはできないが、『国富論』が地理的にも時代的にも多様な国・地域に言及しつつ議論を展開したのに対して、『国富論』が社会に受容されていく過程では、その時代のイギリスの外交関係にとって重要性のある国・地域についてもっばら言及されつつ受容されていった」という仮説を導くことはできよう。

5. 経済学史研究への有効性

Gherity (1993)の著者識別においては、テキストマイニング分析の一部(文章の長さに関する分析)が限定的なヒントを示した。すなわち著者不明の三種類の原稿のうち、一つは多の二つと異なる者が書いたかもしれないこと、および当該原稿の文章の長さはスミスの文章との長さと同様であることを示していること、である。そこから「当該原稿の著者はスミスである」という結論は出てこず、多の分析手法(本稿でいう「経済学史研究の従来の分析手法」と合わせることで仮説の当否に迫る必要があった。

本稿で試みた『国富論』とそれについての書評・書簡類との比較の場合も、テキストマイニング分析は似たような役割と限界を示した。テキストマイニングが明らかにする頻度分析の類似性・相違生に、『国富論』の文脈についての筆者の知識を合わせることで三つの仮説を導き出した。この仮説の当否も、基本的には経済学史研究の従来の研究手法に依拠することではじめて迫ることができるであろう。

しかし看過すべきでないのは、テキストマイニングが、従来の研究手法由来とは異なる性質の事実を我々研究者にもたらしてくれているという点である。前節の(1)～(3)で見た定量的な事実は従来の研究手法では浮かび上がらせにくい。そして用いる文献の量が増えれば増えるほど、このコントラストは鮮明になっていく。

6. おわりに

『国富論』が、当時の社会に受容されていく過程でどのような変容を受けたのか、あるいは受けなかったのか、という課題について、本稿では『国富論』と書評・書簡類とをテキストマイニング(頻度分析)することで以下の四つの顕著な傾向を指摘した。

第一、出現頻度の高い名詞を中心に、『国富論』と書評・書簡類との間には高い類似性が認められた。このことは、[古典] → [書評類] → [大衆] という受容経路を仮定するならば、[書評類] 段階ではかなり正しく [古典] = 『国富論』を伝えていたのではないかという仮説につながる。もっともこの仮説の当否については、伝統的な研究手法も含めて別途、検証しなければならない。全体的には『国富論』と書評・書簡類の間に類似性が認められつつも、

第二、書評・書簡類では『国富論』の貨幣論に関わる単語の出現頻度が低い。このことは『国富論』が社会に受容されていく過程で、貨幣論が置き去りにされていったのではな

いかという仮説につながる。

第三、書評・書簡類では『国富論』と比べて **profit, farmer** の出現頻度が高く、他方で **wages, labourer** や **rent, landlord** の出現頻度が低い。このことは『国富論』が三階級への配分をバランス良く論じていたのに対して、社会に受容されていく過程で、資本家とその利潤に関心が偏重していったのではないかという仮説につながる。

第四、『国富論』が地理的・時代的に多様な国・地域を挙げているのに対して、書評・書簡類では米・独・仏などに偏している。このことは『国富論』の学術的で通時代的な議論が、社会に受容される際には政治的で現代的な関心で受け入れられていったという仮説につながる。

そして重要なことは、テキストマイニングが、経済学史の従来の研究手法とは異なる性質の定量的な事実を提示しており、そこからいくつかの検証に値する仮説を導き出すことが可能となった、ということである。この分析ツールとしての強みは、利用する文献の量が増えれば増えるほど増していく。このことから本稿では、この新しい分析ツールは経済学史研究においても今後広く用いられるべきものであるとした。

テキストマイニングは過度に依拠すべきものではないものの、従来の研究手法を補助するツールとして排除すべきではなく、仮説形成や検証においてツールの一つとして保持しておくべきものであろう。

参 考 文 献

- 金明哲、村上征勝 (2007) 「ランダムフォレスト法による文章の書き手の同定」『統計数理』55 (2), 255-68
- 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」ディスカッション・ペーパー、山形大学人文学部法経政策学科 Discussion Paper Series No. 2012-E02
- 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」弘前大学『人文社会論叢 社会科学篇』31号 51-66
- 樋口耕一 (2014) 『社会調査のための計量テキスト分析——内容分析の継承と発展を目指して』ナカニシヤ出版
- Feldman, Ronen and James Sanger. (2007) *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge: CUP.
- Gherity, James A. (1993) An Early Publication by Adam Smith. *History of Political Economy* 25:2, pp. 251-82
- Google books Ngram Viewer. Retrieved 26th September 2014.
- Miner, Gary *et al.* (2012) *Practical text mining and statistical analysis for non-structured text data applications*. Waltham, MA: Academic Press.
- NaCTeM web page. Retrieved 26th September 2014.

Smith, Adam. WN

Ross, Ian S. ed. (1998) *On the Wealth of Nations: Contemporary Responses to Adam Smith*. Bristol: Thoemmes Press.