質から量に迫る～テキストマイニングと経済学史の方法

小峯敦（龍谷大学）・下平裕之（山形大学）

（1）報告の概要

本報告では、テキストマイニングという定量的解析を含む研究方法が、従来の経済学史研究と接合する可能性・意義・限界を提示する。その手順、および現在までの共同研究の成果を紹介するとともに、新たにベヴァリッジ『失業』（1909）、およびケインズ『確率論』（1921）も対象テキストに含み、新しい知を発掘する。

（2）経済学史の方法論

経済学史研究は伝統的に、対象テキストの厳密な解釈に基づいて、経済学の歴史という観点から、経済（学）的思考の生成・変遷を理論的・思想的に分析してきた。しかし原典解釈を前提とするこの伝統は、対象となる経済（学）から眺めると、二重の意味でますます距離が遠くなってきているように見える。

第1に、経済学史の伝統的な分析手法がむしろ人文科学（特に文学批評）に近いという意味で、社会科学の主要分野と考えられる現代の経済学（理論や計量、あるいは経済史）からの距離が遠い。第2に、社会の量的把握および原子論的把握という特徴を持つ現在の経済学そのものが、質的カテゴリーを多分に含む他の社会科学の諸分野（特に社会学や政治学の質的分析）からの距離が遠い。

経済という現象はそもそも、量的な拡張が可能で成長する（小野塚 2018: 18）ナニモノカをその中核に含むという意味で、社会現象を把握する他の学問領域、例えば法学・政治学・社会学・教育学・心理学・歴史学・民俗学の対象――拡張や成長は考えにくい――とは決定的に異なる。ただし伝統的な経済学の手法は、ペティの量的把握（統計処理を指向）に加えて、スミスの記述的把握（生のテキスト・コンテキストを指向）、リカードの単純因果的把握（因果モデルを指向）を含んでいた。20世紀中葉までに計量経済学が勃興すると、理論モデルのデータ検証という量的把握（データセットの観察）が決定的な重要性を帯びてきた。さらに21世紀には、質的把握の重要な柱であった因果推論が、ランダム化比較実験や自然（疑似）実験という新しい手法によって、量的データの領域にも拡張してきた。

このように経済学が発展する中で、その対象と分析手法がますます量的データの解析という自然科学に近づき、人文科学のみならず他の社会科学（そして文献解釈を主とする経済学史）からも隔絶した領域となった。他方、自然科学の中には「量から質に迫る」として、人間個体の感性を解析する動きも進んでいる。例えば物理的な現象（可視領域の電磁波、空気の振動、粘膜の化学的変化）を電気信号に変えて神経に伝達される仕組みが感覚（特に五感）であるが、脳科学その他の発展によって、感性の画像的・数量的把握が急速に進んでいる。

ただし、このような単純な感性を超える《高次感性》（村井編 2014: 3）が人間の社会を支

えている。芸術家のモチーフ、起業家のアイデア、政治家の決断力などは、五感に比べてさらにつかみどころがない。これらは複合的な感性の集まりであり、社会的な背景や過去の知識に強く影響を受ける。経済学史の主な対象を、経済学者の内部論理とその世界観・洞察力に分ければ、後者が《高度感性》に対応する。

本報告は以上の基づき、計量テキスト分析（テキストマイニングの一種）を２つの視点を加えて実践することで、従来の経済学史の方法論では欠落していた視点を補完したい。第１に、Goertz & Mahoney (2012: 4／訳 5)が除外した《解釈アプローチ》の妥当性について、定性・定量という論争を見据えて論じる。第２に、経済学方法論を経済学と科学論の間の相互浸透と定義した Hands (2001: 7／訳 7)を典型に、現在の経済学の方法論は主に科学哲学や認識論から考察されてきたが、ここではむしろ隣接分野である現在の社会学や政治科学において、統計処理に対抗して急速に整備されてきた視点（質的カテゴリー）を取り上げる。

（3）テキストマイニングと計量テキスト分析

原典・テキストに内在する論理や外在する価値観をより客観的に把握する試みの１つが、テキストマイニングである。従来、この手法は「テキストからの知識の発見」（喜田 2008: 27）という説明を典型に、明確な定義が困難であった。近年では「テキストデータを、言語処理技術を用いて構造化データ・変数に変換し、それをもとに知識発見、仮説発見および仮説検証を行う手法」（喜田 2018: 8）という具合に、自然言語処理（形態素解析→構文解析→評判分析など高度な解析）とデータマイニング（頻度集計→統計処理→視角化）を混合させる方法として認知されている。原典を量的（大量データの機械的処理）および質的（少数データの符号的処理）に数値化・視角化することで、テキストの表層に隠れた構造・意味・テーマを発掘する手法である。

ただし本報告ではさらに一歩進めて、小林（2017a: 19）が提唱するように、自然言語処理とデータマイニングをこの順で適用し、仮説の設定と検証の相互交流を明示化させる手法を採用しよう。この明示化について、ソフト（KH Coder）の開発者・樋口（2014: i）は「計量テキスト分析」という名称を与え、大量データを自動処理する第一段階、質的カテゴリー（可視化された分析者の視点）を加味する第二段階という二段階解析を勧めている。

（4） ４段階からなるテキストの量的・質的把握

テキストという自然言語をどのように量的に把握するのか。テキストマイニングの過程を次の４段階に分けよう。①構想と前処理：テキストマイニングは自動的に分析が開始されるわけではない。《やわらかな作業仮説》と呼ぶべき、分析の目標がまず設定されていなければならない。これは先行研究の知見に基づく。さらに、前処理として、分析ソフトに読み込めるためにデータの収集・洗浄が必要である。データ収集としては、対象となる原典（活字や手書きのテキスト）はデジタル化され、文字認識（OCR）されなければならない。その上で、デ

ータ洗浄として、構造化されていないデータ――脚注や表や見出し語やページ数が入ったままのデキスト――を様々なソフトや目視によって整理する作業――一語一語が検索可能な形となり、数値や文字を項目とする行列の表という電子データを作成すること――に進む。この作業には非常に多くの時間と費用がかかる。②自然言語処理：洗浄されたデータが手に入ったら、形態素解析（単語分割）、係り受け（構文）解析、（正負の）感情分析・（潜在的な）意味解析などの高度な解析、という 3 つが機械的に判定される。ただし、この段階でも品詞の指定や除外語・特定語の指定など、研究者の意図も小さいが混入している。

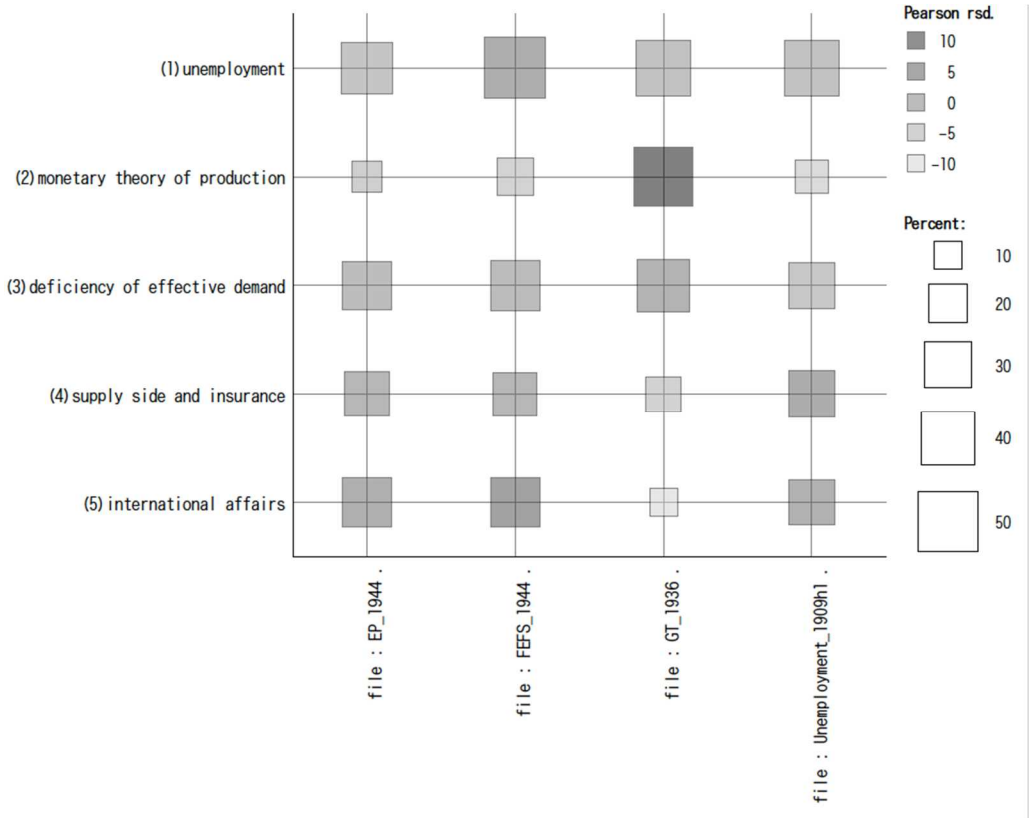
③データサイエンス：頻度集計、様々な多変量解析、視覚化という三段階を経る。単純だが最重要の分析は単語の出現頻度である。ここで単語の量的な把握に警戒感があるのは当然である。しかし、スミスの《見えざる手》を例に取れば、a)出現回数は『国富論』で本当に一回だけなのか、b)出現頻度が極端に低いにも関わらず、なぜこの言葉がキーワードとなっていくのか、などを当該テキストや周辺のテキスト（コーパス）を解析することで、ある主要単語の中心性や媒介性が明らかになる。もう 1 つの多変量解析の基本は、共起である。共起とは、ある文脈（文・段落・節・章・部など、その範囲を自由に設定する）において、ある単語とある単語がどのくらい同時に出現するかという指標である。各単語はそれぞれの共起回数を要素とする《単語ベクトル》を有しており、双方のベクトルのなす角度を調べることによって《類似度》を定義できる（同じ文脈に出現する単語は類似した意味を持つという「分布仮説」に基づいている）。これがテキストという質から類似度・距離という量に変換される仕組みである。こうした統計量に基づいて、各単語・各単元の連環は量的に把握され、視角的に表現される。ここまでの段階でも、分析範囲や指標選定において、様々な分析者の観点は入っている。④最後に、いったん機械的な分析が終了した後に、改めて分析者の独自な観点を取り入れた《コーディング・ルール》（検証可能な観点からテキスト全体を分類する工夫）に基づいてカテゴリーを新設する。この段階でいったん切断された単語が、有意味な意味内容として再構成される。仮説を組み立て、自動処理と質的カテゴリーという二段階で検証し、さらなる洗練された仮説を組み立てることになる。

（5）事例研究 1 ：ベヴァリッジ『自由社会における完全雇用』

本報告では、現在までの共同研究の成果を要約し、加えて国内外の類似研究についても例示する。まず事例 1 として、ベヴァリッジ『自由社会における完全雇用』（1944）に、どの程度、ケインズ『一般理論』（1936）の要素があるか、ベヴァリッジ『失業』（1909）と政府白書『雇用政策』（1944）を補助線に用いて確定しよう。

重要な複合語を自動的に抽出する TermExtract を用いたところ、『一般理論』の最重要語が rate of interest、次が marginal efficiency of capital、三番目が quantity of money だと判明した。他方、『自由社会』では、完全雇用・国際貿易・労働市場という具合に、従来の研究を裏付ける重要語群が上位に来た。そのため、改めて独自に 5 つの範疇（下図の縦軸）を設定

し、それぞれに含まれる単語（下の左表）を確定した。この範疇に含まれる単語が、４つの作品について何%含まれるかを視覚化した結果が次のバブル図である。



ここから、「生産の貨幣理論」範疇が『一般理論』のみで特徴的であり、「供給側と保険」・「国際関係」範疇は『一般理論』のみで特徴的ではない、ことがわかる。さらに、「失業」範疇は４つともかなり似通っており、特に「有効需要の不足」範疇に関しては、４つの作品でほぼ同じ割合となっている（３作品では、カイ二乗検定で有意差が検出できなかった)。

*(1)unemployment
employment or unemployment or output or income

*(2)monetary theory of production
rate or interest or marginal or efficiency or quantity or money or monetary

*(3)deficiency of effective demand
fluctuation or demand or chronic or deficient or unsatisfied or deficit or lack or shortage or investment or outlay or comsumption or expenditure or aggregate

*(4)supply side and insurance
supply or exchange or insurance or location or irreducible or minimum or maladjustment or industry

*(5)international affairs
international or multilateral or balance or terms or trade or foreign or peace or war or import or export

	A	B	C	D	E	F
1	Noun		ProperNoun		Adj	
2	probability	1405	O	394	other	399
3	case	537	Bernoulli	122	same	283
4	argument	467	PROBABILIT	115	probable	267
5	h	456	Principle	111	such	258
6	instance	440	P	106	true	249
7	knowledge	437	B	91	certain	245
8	number	412	Theorem	89	more	211
9	proposition	381	S	88	possible	209
10	value	315	Laplace	86	first	197
11	b	306	E	75	particular	172
12	p	296	Q	66	general	170
13	series	287	Indifference	65	independent	161
14	evidence	284	Mr	59	statistical	151
15	method	275	Boole	56	logical	147
16	conclusion	272	Induction	56	different	141

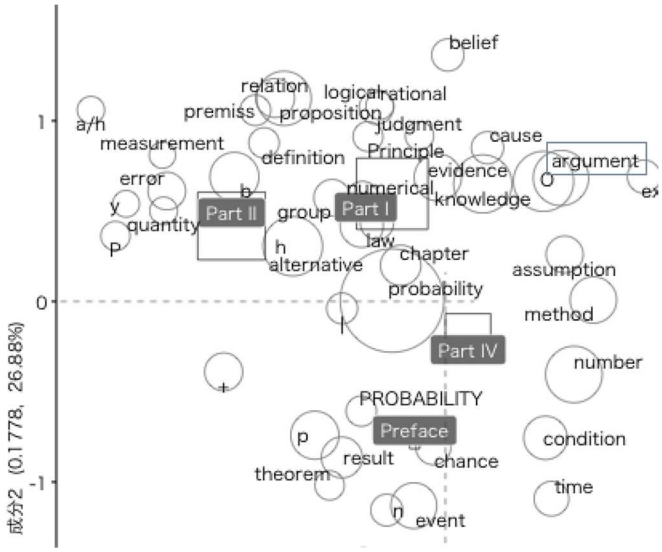
コーディングルール（左）と『確率論』の品詞別頻度（右）

以上の計量テキスト分析から、①『一般理論』では「生産の貨幣理論」が他とは異なる特徴を持つこと、②有効需要の不足、失業論というテーマは『自由社会』でも『雇用政策』でも引き継いでいること、③供給側・保険・国際関係という『失業』のテーマは『一般理論』には薄

く、ベヴァリッジ独自の初期思想として持続していること、などが確証できた。

（6）事例研究 2：ケインズ『確率論』

ケインズ『確率論』を分析するために、ナイト『危険・資産および不確実性』、ケインズ『一般理論』第 12 章、ケインズ「雇用の一般理論」1937 との対照を試みた。上記の右表は、『確率論』の品詞別頻度であり、下図は対応分析である。この図では原点を特徴がない（＝普遍的な）点と見なし、そこから縦軸にも横軸にも遠ざかるほど、ある範疇・単位ごとの特徴が明示化されている。図では、『確率論』の 5 つの部（章が集積された単位）の特徴を抽出している。この方法では他の著作とのトピックの差を明示することができる。他にも本報告では、階層的クラスター分析、共起ネットワーク分析、KWIC コンコーダンスなどを例示する。



『確率論』各部の対応分析（部分）

（7）おわりに～経済学史をひらく

最後に、「質から量へ迫る」というプロジェクトを通じ、テキストマイニングが経済学史研究を、二重の意味で《ひらく》可能性についても言及する。第 1 に、経済学史の研究者がより深く自然科学的な分析手法を実際に運用すれば、経済理論家だけでなく、多くの科学者の思考法をより深く理解できる可能性を秘めている。その理解は単に受容するという意味ではなく、量的把握に潜む問題点の指摘など、批評的な眼も涵養されることを含む。第 2 に、特に類似の方法論を経済学史研究に組み込むことができれば、社会思想史や政治哲学などに加え、他分野（社会学、政治科学、統計学、情報処理学、計量言語学、データサイエンス等）の研究者との協働・分業を通じて、より説得的な、より新しい知見をもたらす触媒ともなりえる。

参考文献（主要なものは次の文献表を参照のこと）

小峯敦・下平裕之（2017）「ベヴァリッジ『自由社会における完全雇用』のケインズ的要素～テキストマイニングを加味した量的・質的分析～」*Discussion Paper Series*, Faculty of Economics, Ryukoku University, No.17-01: 1-70, September 2017.

村井源編（2014）『量から質に迫る～人間の複雑な感性をいかに「計る」か』新曜社。