

19世紀から20世紀の書籍ビッグデータ分析に見られる近現代美術運動の興亡の理解

Understanding the rise and fall of Modern and Contemporary art movements through Big data analysis of the books published in 19th and 20th century

沖 啓介
OKI Keisuke

はじめに

美術運動の発生、隆盛、衰退は、従来は美術史研究で研究されてきており、その蓄積は重要であるとは言うまでもない。

他方で、美術史研究の方法は、情報技術の発展によって、これまでにない視点からの方法が可能になりつつあり、これは美術史に限らず、歴史学全体に波及していくものと考えられる。

人間の諸活動も生物学的な営みであり、そこには生命活動に見られるような誕生、成長、衰退までを含むプロセスを見ることができるはずである。

デジタル技術によって広範にデータを集積することが可能になるまでは、社会的現象を生物学的な現象のように読み解く技術は存在せず、人々は記憶を掘り起こすものとして、書籍や文献、またその集積所である図書館やデジタルメディアのような記憶装置を通してほんやりとその興亡を把握してきた。だがそこでは動的に把握するビジョンは持ち難い。

しかし近年、「ビッグデータ」という巨大データがあらゆる領域で蓄積されるようになり、社会における人間の諸活動を示すデータを得ることが可能になっている。

本研究では、19世紀から20世紀に出版された書籍のデータから、近代、現代の美術運動の勃興から現在まで、その運動がどのように推移してきたかに注目することによって、この時期の「美術史」(あるいは芸術史)の断面を観察するものであり、美術運動に関する書誌学的なあり方にも今後の可能性を示すものとした。

1. 印象派に見る美術運動の発生パターン

「美術運動」は、20世紀の大半を通じて、美術史の歴史的な分類でも重要な意味を持っている。(この芸術的な創造活動に「名称」が与えられ、分類されることは、美術にとどまらず、文学、音楽、デザイン、工芸、映画、建築までにも及んでいる。)

それらの「発生」を巡る状況は多様で、「印象派」のように表現傾向の似た芸術家集団が登場し、それが評論家らの批評によって、例えば「印象にすぎない」と言われたのを逆手

にとって、生起している現象に集合論的に命名したものであったり、「超現実主義」や「イタリア未来派」のように文学表現とシンクロすることで思想性を明らかにして「宣言」されたものもある。

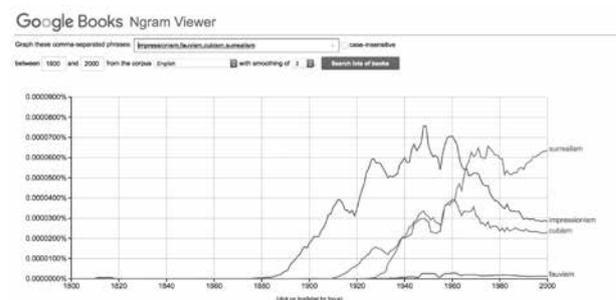


図1

印象派の発生については、すでに周知のことであるが、その経緯は次のようなものである。

- 19世紀後半にパリで出現した少数の画家の集まり
- 保守的な美術界からの批判にさらされる
- 連続して展覧会を開催し1870年代から1880年代にかけて突出してくる
- 運動の名称の由来はメンバーの一人クロード・モネの絵画作品「印象・日の出」による
- 批評家ルイ・ルロワによって風刺新聞「ル・シャリバリ」上で「印象派の展覧会」(あるいは「印象主義者の展覧会」)という文章で批評される

「印象派」(Impressionism (英) impressionnisme (仏))という名称は、自身も画家でもある美術批評家に揶揄されてついた名前だが、それがこの芸術運動の本質的な内容の一面を表しており、その時代のラディカルな表現として広まっていた。

図1は、印象派 impressionism、キュビズム cubism、野獣派 fauvism、超現実主義 surrealism という19世紀末から20世紀初頭に登場した美術運動の名称が、英語で出版された書籍の総数のなかで取り上げられた「割合」を示したものである。

このデータを見ると19世紀、20世紀初頭の美術運動の取り上げられかたは、20世紀中期に大方のピークを迎えている。しかし超現実主義はまた他とは違った様子を見せている。(それは超現実主義は、美術運動だけでなく、文学運動でもあり、映画でもあるからだろう。)

この統計は、Google社のGoogle Books Ngram Viewerプログラム (<https://books.google.com/ngrams>) を使ったもので、同社が19世紀から現在までの英文書籍のデジタルデータ化を行なっているものから検索したものである。(これまで英語のデジタルデータが中心であったが、他のヨーロッパ言語や中国語も加わっている。)

図2は、同プログラムのフランス語書籍を同様に解析したものである。

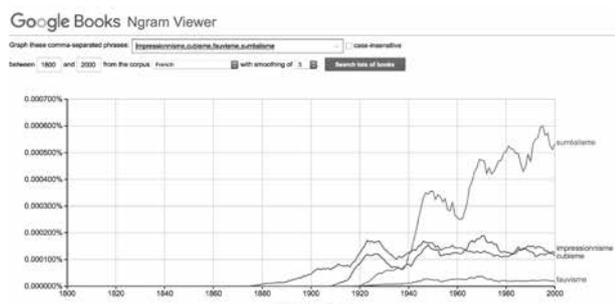


図2

英語とフランス語のデータからそれぞれの美術潮流の発生のおおよその時期を特定することができる。

英語で調べたものが図3であり、さらにフランス語で「印象派」を意味する *impressionnisme* の頻出度をフランス語の書籍データから調べてみたのが図4である。

これらのデータはもちろんGoogle社が収集している書籍データに負っている。デジタル書籍以前の、紙に印刷された文字をOCR(光学式読み取り)技術でデジタル化した結果である。そういう意味で、書籍のOCR化という作業に負っている点を考慮すると統計データの精度には不確かなものもあるかもしれない。

しかし厳密な精度にこだわらないのであれば、大まかには大体の傾向を読み取れるのではないかと考えられる。

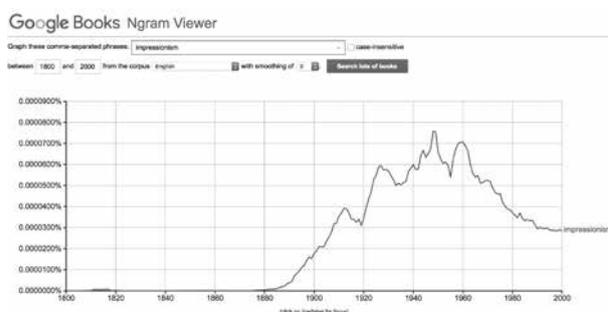


図3

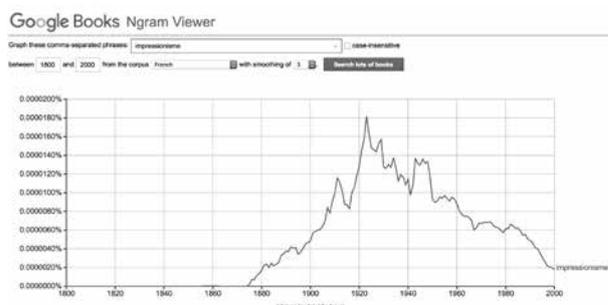


図4

これらのデータから印象派という美術運動が発生した時期と、それが120年ほどの時間を経てどのような概念的なダイナミズムあるいは書誌学的な推移をたどっているのかがわかるのと同時に、美術表現の概念が国や文化を超えてどのように推移するのとも見えてくる。

データには不完全さやノイズがつきものであるのだが、この書籍データでも同様である。例えば「印象」を意味する *impression* という単語はごく一般的であり、そこから派生した *impressionism* という単語が、美術の文脈とは別の分野で出てくることもありうる。

科学的な研究、調査では、このようなノイズや例外はむしろ無視して、論理的な帰結に結びつく「法則」を現象から抽出していくものかもしれない。

この場合も、もちろんノイズを排して、美術運動の「印象派 = *impressionism*」に集中するというのも方法論的に間違いではない。

つまり、「印象にすぎない」という批評があって、その言葉尻を捉えて自分たちの表現の概念と名称に集約させ、しかもその名称を用いることによって社会的に認知されるようになったと理解すると、一見すると無関係なノイズ部分にも文化的な背景として関連性を見ることができないのではないだろうか。

文化というのは単純なパラメータではなく複雑なパラメータを持っている。このことは美術運動の名称が、他の分野にも影響を与えるようになる20世紀後半以降の文化現象とも似てくる。

例えば「未来派」を意味する「futurism」という単語は、イタリアやロシアの幅広い表現領域にまたがる前衛芸術運動「未来派」とはまったく別に、現在では、未来を意味する思考や大衆文化に見られるようになっていて、すでに検索結果からだけではこの前衛芸術運動だけを抽出して捉えられない。

つまりここから先は、その単語を文脈から把握しなければならないのだ。

同様なことは20世紀後半の美術表現である「ミニマリズム」(minimalism)にも起こりうる。つまりfutureとかminimalという一般的にも多用される単語の場合、それはまったく異なる分野で、また異なる意味で使用される。

また同じく20世紀初頭の前衛芸術運動のダダイズム(dadaあるいはdadaism)の場合も同様である。dada命名の由来ははっきりとしないが一説によるとオーストリア人アーティストのリヒャルト・ヒュルゼンベック(Richard Huelsenbeck)が無作為にナイフを辞書に差し込んでたまたま出てきたフランス語の「幼児用馬乗り遊具」の名前を使用したと言われている。そういう特殊な命名方法をとったわけだが、いまや「dada」は化粧品やファッションなどを中心に様々な製品の名前として使われている。

このようなことにあまり影響されにくいと考えられる「超現実主義」(surrealism)のような例もあるが、それでも20世紀に勃興した商業主義のダイナミズムには抗しきれないだろう。(また逆の視点からすれば、文化は多かれ少なかれ、いつでも時代の経済、商業、大衆文化、メディアとの関係のなかで成立しているということだ。)

2. 概念伝播の時間

書籍データから個別の美術運動の盛衰を見ることも、ここでの興味深いことである。例えば、上記の「印象派」についての引用の増減数もその一つである。図3、図4は、それぞれ英語、フランス語の書籍データである。分母となる発行部数が異なるが、それぞれの言語での「印象派」という単語が出てくる数が、全出版物の文字部分の量からの割合で表示される。そういう意味では社会的な認知度などはわからないが、出版物にみられる「印象派」という言葉への関心の度合

いの傾向が表れているだろう。

ここでも当然の事ながらノイズは存在するだろうが、それぞれの言語での「印象派」と見られる引用された単語の量が出てくる。それにより英語とフランス語のそれぞれの言語での「印象派」への「関心度」の傾向が見られる。これらを概観すると、フランス語での印象派という言葉の量は、1920年代をピークにして下降傾向にある。一方の英語では、1950年代あたりをピークにして下方へと降下している。

またリバイバルとして一度下降した傾向が再度上昇することもある。それはフランス語では、1940年代を通じて上昇して同じ水準が続き、一方の英語では1950年代半ばから1960年代半ばまでの間で1960年あたりをピークに弧を描いている。

これらの傾向を、印象派の作品の紹介の頻度や、印象派の概念の伝播という観点から、どれだけ実態をつかめるかは、データ細部の読み込みが必要であるが、少なくとも何かしらのタイムラグがあることだけは理解できるだろう。

また文化の伝播という点では同時代の短い時間に、国を越えてどのように美術運動が伝わっているのかも見て取る事ができる。

英語版の印象派データから初期の書籍の次の内容を見ることができる。

1893年に英国で出版された“Art Journal”第55号では、103ページにある“Some Remarks on Impressionism”(印象派についての感想)[註1]という文章を掲載してとても興味深い。英国のエディンバラにあるスコットランド国立美術館(Royal Scottish Academy)の館長ジョージ・ライド卿(Sir George Reid)が、当時有力であったロンドンの新聞“The Westminster Gazette”で、「印象主義者運動」について述べていることを紹介し、その文面では、彼がスコットランドでは印象派の影響を受けてその様式を単に真似て活動する者がいることを嘆いている。また「印象派は、我々の知性に訴えかける極めて高度な芸術様式である」というプラザーウィック博士(Dr. Blatherwick)なる人物の評価コメントも紹介している。この文のまとめでは、英国の印象派はフランスの模倣(imitation)ではないのだというライド卿の言葉を引用したりしている。

ここに見られる構図は、西洋美術が持ち込まれた明治期の日本の状況とも似ているだろうが、「輸入文化」のようにして取り込まれる現代の文化にも共通するものがあるだろう。

この印象派の英国での受け止められ方は、同時代の美術様

式の伝播のあり方を示す例として興味深いものである。

3. 概念の「強度」

前の章で「概念伝播の時間」を紹介したが、概念伝播には、それぞれの「強度」があるとも考えられる。これを物理学的な表現の「強度」と呼ぶのが適切かどうかは論議が必要だが、上昇する増進と、逆に下降する減衰を見ることができる。

この点で興味深いのは、図1と図2に見られる「印象派」(Impressionism, Impressionnisme)、「立体派」(Cubism, Cubisme)、「野獣派」(Fauvism, Fauvisme)、「超現実主義」(Surrealism, Surréalisme)という近代美術の4つの潮流のデータである。

まず英語、フランス語の書籍データに共通するのは、「超現実主義」が他の美術潮流を圧倒してトップであることだ。

このことをどのように解釈するかは様々に考えられるだろうが、超現実主義がもっとも「現在の」、つまり現在にも訴求しやすいものがあると考えられるかもしれない。また先述したように超現実主義は多様なメディア表現性を持っていることも挙げられるだろう。一方で印象派は、そのものへの言及が減っているのは、図3、図4でもわかるのだが、フランス語では、立体派と同程度であるかのように見える。

英語で見れば、印象派、立体派がどちらもピークを過ぎて下降しているのだが、一方で超現実主義が、勢力を保ってトップという、その文化的背景は考察に値するものだろう。

このような現象がどうして起こるのかを、解明するには時間が必要であるだろうが、それを認識することができたのは、書籍データの収集が行われるようになったからである。

4. 検索エンジンと N-gram

Google Ngram Viewer のシステムの基礎となっている「N-gram」(Ngram)という手法は以下のようなものである。

これはインターネットの「検索エンジン」で使われている技術である。多くの検索システムでは、キーワードとページでの位置の組み合わせを用いるのが一般的である。これを「転置インデックス」(Inverted index)といい、全文検索を行う対象となる文書群から単語の位置情報を格納するための索引構造をいう。

その転置インデックスのキーの切り出しを、辞書や構文解析に基づくのではなく、単に一定の文字数で切り出した語を

入れることで作る方式のことを N-gram という。[註2]

N-gram

- 検索対象のテキストを N 文字単位の文字列片に分解する
- 分解した文字列片を見出し語として転置インデックスを作成する
- 検索語を N 文字単位の文字列片に分けて検索を行う
- 文字列の出現位置情報を利用すれば、漏れない完全一致の検索が可能

ちなみに検索エンジンでは、「形態素解析」という方法も用いられている。

形態素解析

- 検索対象のテキストを形態素解析を行い分かち書きを行う
- 分かち書きした単位を見出し語として転置インデックスを作成する
- 転置インデックスを元に検索を行う

違いは、「転置インデックスの見出し語をどのように作るか」というプロセスが異なり、形態素解析は構文解析を行って分かち書きを行うのに対して、N-gram では機械的に N 文字の文字列片に分けるという処理となる。N-gram による見出し語の切り出しは、形態素解析のための文法解析を伴わないため、特定の自然言語に依存しないという特徴がある。

この「転置インデックス」(inverted index)とは、全文検索を行う対象となる文書群から単語の位置情報を格納するための索引構造のことである。

例えば、 T_0 ="it is what it is"、 T_1 ="what is it"、 T_2 ="it is a banana"という3つの文章があったとする。この場合の転置インデックスは以下ようになる。[註2]

```
"a":      {2}
"banana": {2}
"is":     {0, 1, 2}
"it":     {0, 1, 2}
"what":   {0, 1}
```

{2}とは T_2 の文章のことであり、{0, 1, 2}とは T_0 、 T_1 、 T_2 の文章にそれぞれの単語が含まれていることを示している。"what", "is"そして"it"といった語に対して単語検索をかけてみると、次のような集合が得られる。

$\{0, 1\} \cap \{0, 1, 2\} \cap \{0, 1, 2\} = \{0, 1\}$

同じテキストから文書番号と単語の出現位置まで含んだ完全な転置インデックスをつくと次のようになる。

"a": $\{(2, 2)\}$

"banana": $\{(2, 3)\}$

"is": $\{(0, 1), (0, 4), (1, 1), (2, 1)\}$

"it": $\{(0, 0), (0, 3), (1, 2), (2, 0)\}$

"what": $\{(0, 2), (1, 0)\}$

配列を用いているので0を基点にするので、つまり

"a": $\{(2, 2)\}$

という場合は、 T_2 の3番目の位置に"a"があることであり、

"is": $\{(0, 1), (0, 4), (1, 1), (2, 1)\}$

の場合では、 T_0 の2番目と5番目の位置に、 T_1 の2番目、 T_2 の2番目の位置に"is"があることである。

ここで"what is it"というフレーズを検索すると、文 T_0 と文 T_1 がヒットする。だが、この三つの単語がフレーズとして現れるのは、文 T_1 だけである。

この出現位置情報などを含めて「コーパス」を作っていく。コーパス(英: corpus)は、言語学において、自然言語処理の研究に用いるため、自然言語の文章を構造化し大規模に集積したものである。

検索エンジンはこのような構造で自然言語の処理を行っている。日本語でも古文や和歌などのテキスト分析がこの方法で行なわれている。

またN-gramは、検索エンジンでの自然言語処理以外に、DNAを構成しているヌクレオチドの塩基配列の決定(gene sequencing)、またはタンパク質のポリペプチド鎖内のアミノ酸配列の決定(protein sequencing)を行う手続きなどにも利用されている。

5. 書籍データ検索の結果とカルチャロミクス

Googleの創業者のラリー・ページ("Larry" Page)とセルゲイ・ブリン(Sergey Brin)は、"The Anatomy of a Large-Scale Hypertextual Web Search Engine"(大規模なハイパーテキスト的なウェブ検索エンジンに関する分析)という論文を共同執筆をした。もともと彼らが学生時代にやろうとしていたプロジェクトに、彼らが学んでいたスタンフォード大学の図書館所蔵の書籍をデジタル化して公開することがあった。2004年12月には、オンラインで公開している。[註4]

ページらの試みは、大学図書館の蔵書にとどまらず、全て

の書籍のデジタル化を目指すようになっていく。その結果が、Google Booksという膨大な書籍データベースとなった。

書籍が著作物であるので、著作権保有者側からすると、書籍の中身をデジタル化してGoogle Booksが公開することには反発があった。しかしそれにもまして、書籍データの(しかも膨大な量の書籍のデータ、つまり書籍のビッグデータによって)検索が行えることによって従来では見ることができなかった文化的な大きな「視座」が生まれていることの意義は大きい。

Culturomics(カルチャロミクス)という分野がそれである。日本語では漢語化された訳語がまだ存在しないのだが、中国語では「文化組学」という中国語としての適訳がなされている。つまり“文化”(culture)と“組学”(omics)が組みあわされたものである。

ここでのポイントは、「オーミクス」(-omics)という接尾辞部分にある。これは非公式なものではあるが生物学での方法で、「研究対象+omics」という名称を持つ生物学の研究分野の命名方法に由来している。例えば、遺伝子研究であれば、genomics(genomics = gene + omics)であり、生物体内に存在する一次転写産物(すなわちmRNA)の場合は、トランスクリプトミクス(transcriptomics = transcript + omics)となったりする。

そういう意味で、カルチャロミクスは、人間文化についての生物学的な観点を含んだ研究分野とも考えられるだろう。この概念は、エレット・エイデンとジャン＝パティースト・ミシェルらの共同執筆による"Quantitative Analysis of Culture Using Millions of Digitized Books"(大量のデジタル化書籍利用による定量解析)という2011年の論文で初出になったものである。[註5]

彼らは、言語利用に於ける文化パターンを分析するために、書籍ビッグデータであるGoogle Booksを、N-gramで分析するGoogle Ngram Viewerにも関わった。

彼らは、「デジタルメディアを利用すれば、情報の内容の変更や処理を確実にこなせるようになる。十分な量のデジタル化された記録と、それを処理するに足るコンピューター能力があれば、人間の文化をこれまでとは異なる新たな視点から眺められるようになり、われわれの世界とわれわれが置かれている状況の理解の仕方にきわめて大きな影響を及ぼす可能性がある。」[註6](Uncharted: Big Data as a Lens on Human Culture 邦題「カルチャロミクス:文化をビッグデータで計測する」第1章 歴史を見通す見通す新しい眼鏡)と述べている。まさに英語の原題の「人類文化のレンズ

としてのビッグデータ」であり、まさに「未踏」(uncharted)の地帯である。

もちろん収集されている19世紀からの書籍データにバイアスが無いとは言えないし、また印刷物をOCRでスキャンする時に生じるエラー(書体の読み間違い。例えば古い活字ではfとsが酷似していて、読み間違いがあることが指摘されている。)などは多々あると見られている。[註7]

しかし、それらはGoogle Ngram Viewerのデータや技術の不完全さであるだけで、書籍のビッグデータ分析という手法そのものは、画期的な技術だと言えるだろう。

このような動向の影響を受けたカルチャロミクス・プロジェクトもその後も多く登場している。[註8]

「Culturomics 2.0」というタイトルで、カレブ・リータル(Kalev H. Leetaru)

が行った研究プロジェクトでは印刷と放送メディア(テレビやラジオの放送原稿)を使っている。この研究では、2011年の「アラブの春」の兆しと、オサマ・ビンラディンの最後の居所を約200キロメートル以内で突き止めたことで知られている。

物理学者のアレクサンダー・ピーターセン(Alexander M. Petersen)らは2012年のNature誌での論文“Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death”で、言語の誕生と死滅の割合を調べて、その死滅の増加と誕生の鈍化の劇的な変化を指摘している。

「世論における傾向の変化」をソーシャルネットワーク・システムのtwitterでの投稿から分析したもので、T. Lansdall-Welfareらは、980万のユーザーによる8400万の「ツイート」で英国での政府支出削減の発表から3ヶ月を調査し、大衆心理の変化を調べている。また彼らは同様に、福島原発事故の前後の5年間に500万にのぼるニュース記事を収集するなかで、原子力に対する急激な心理変化を調べている。

このようなカルチャロミクスのアプローチは、多様なメディアを対象にした研究で応用されている。

6. ビッグデータから抽出する美術の要素

「美術作品」とその作り手である「アーティスト」、そして「美術批評」が、美術とその歴史を形作る構成要素であると考えるのが、伝統的な思考法であるだろう。作品論や芸術家論が批評という形式で整理されて美術史を構成してきた。もちろんこの方法は、文化的にも有効な方法として今後も残っ

ていくだろう。

もう一方で考えられる要素は、アートが持つ「ダイナミズム」とも言うべきもので、そういう動的な要素は、歴史のなかでは瞬間的なものであったり、発現するのに時間があったり多様な存在の仕方をしている。

アートが社会、歴史などの観点から成立するには、作り手(アーティスト)だけでなく受け手(観客など)が必要であり、受け手側の「動向」「反応」を把握する必要がある。しかしこれまでは、受け手の動向や反応を把握する決め手となるものはなかった。

「歴史の流れを定量的に捉える」[註9]のが、カルチャロミクスの目的であるとする、その指標として書籍データを使用し、アートあるいはアーティストのあり方を見出していくことは、意義のあることではないだろうか。

7. 「名声」とは何か

「名声」というのは、芸術表現にはつきものである。芸術の歴史で考えるとまさに「名声」こそが、不変の価値を決定する要素であるだろう。

図5は、20世紀に美術界大きな影響を与えたアーティストとして、パブロ・ピカソ Pablo Picasso、サルバドル・ダリ Salvador Dali、アンディ・ウォーホル Andy Warholを英語の書籍を対象にして、Google Ngram Viewerにかけてみたものである。

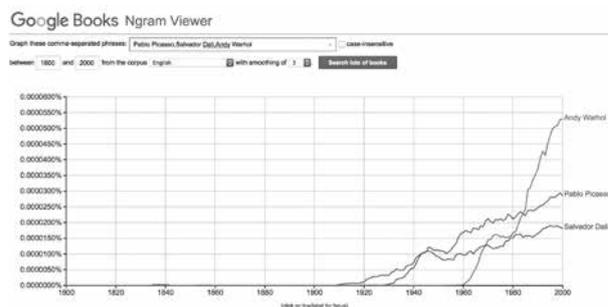


図5

そもそも「名声」とは何であるのかという定義が必要であるかもしれない。書籍により多くの名前が引用されるのが「名声」であるかどうかは論議の余地はあるかもしれないが、一般的には引用の数は一つの「尺度」と考えられる。

ピカソの生涯は1881年10月25日 - 1973年4月8日、ダリの生涯は1904年5月11日 - 1989年1月23日、ウォーホ

ルの生涯は1928年8月6日 - 1987年2月22日である。厳密にはピカソは19世紀の生まれであり、他の二人は20世紀に生死している。

ここで印象派画家のピエール＝オーギュスト・ルノワール Pierre-Auguste Renoir をあえて加えると図6の結果が得られる。ルノワールの生涯は1841年2月25日 - 1919年12月3日であり、印象派が突出してくるのが1870年 - 1880年あたりであり、ほぼ19世紀のアーティストである。

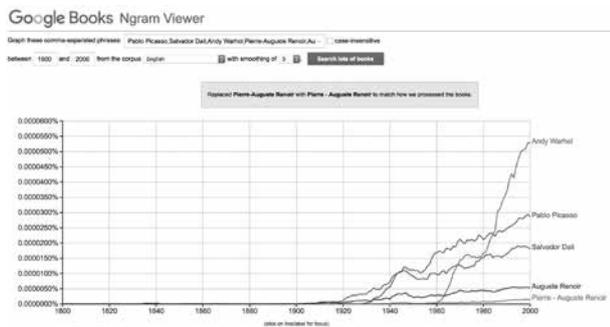


図6

ルノワールの場合、名前の表記をどうするかによって結果が変わってくる。ここでは Pierre-Auguste Renoir と Auguste Renoir の二つの表記で調べてみた。(その結果では、後者の方が多い。表記については Google Ngram Viewer の検索方式に負っている。)

先にも述べたようにこれらの数値は年ごとの「出版された書籍の総数のなかで取り上げられた割合」である。総じて言えることは、アートのは題は出版されて採集された全体の書籍のなかで、増加傾向にあるのが見て取れる。そしてルノワールが引用されることは、このアーティストたちと比較すると少ない。

7. 変容する文脈を捉える

「名声」という言葉を前章では使ったが、これはむしろ「認知度」として用いた方がいいのかもしれない。

人物として書籍などで名前が取り上げられる割合は、アンディ・ウォーホルがルノワールよりもはるかに多い。このことをどのように分析するかは、これらのアーティストが取り上げられている書籍の内容を精査しなければならないだろう。

しかし概観として私見を述べるならば、ウォーホルは極めて語りやすいアーティストであると言えるだろう。それはメ

ディアの発達が19世紀とは比較にならないほど拡大しているということにも関連している。

ウォーホルは、語られるべきことが多い。しかし歴史に名を残す芸術家というのは、歴史を越えて語られるべきものに満ちているとも考えられる。

そこで時代は異なるが、芸術家として数奇な運命を辿った音楽家のアマデウス・モーツァルト Amadeus Mozart と、画家であり発明家、科学者でもあり、また小説「ダヴィンチコード」のような大ヒットした大衆小説にも登場するレオナルド・ダヴィンチ Leonardo da Vinci を加えて比較してみたのが図7である。

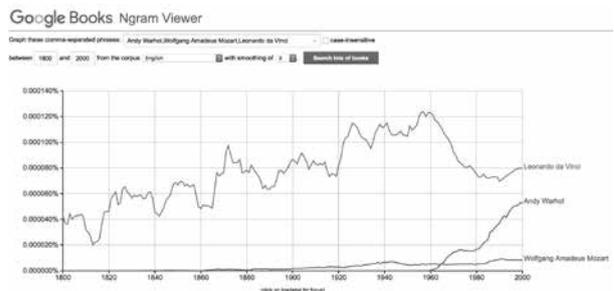


図7

この図から何を読み取るのかを、断定的に言うことはできないが、ダヴィンチが語られることがいかに多いのかということがわかる。一方、奇妙にも1960年をピークに彼を語ることが減少していて、ウォーホルと対称を成しているようにも見える。ここに何らかの因果関係があるのかは現状ではわからない。

8. 書籍ビッグデータと統計的自然言語処理

これまで見てきたデータとその結果というのは、美術運動の発生やアーティストのデビュー以降の時間にそってどのような反応が持たれてきたのか、あるいは認知されているかを知る手がかりになるだろう。それだけでも従来にはなかった方法であり、その意義は大きい。

しかし、それらの美術運動やアーティストがどのように書籍の文中で捉えられているのかは、その文章を読んで理解するしかない。それも意義深いことであり、こつこつと読み比較することで見えてくるものは大いにあるだろう。

だが、この先は、美術運動やアーティストがどのような文脈で語られているのかが重要になってくる。単純な分類であるが、例えば肯定的であるかあるいは否定的な文脈で捉えら

れているかというのも、初期の分類の例であるだろう。そしてさらに表現技法、思考、思想、社会性、交友関係、家庭環境、経済、幼年時の様子、教育、病歴・・・などに関して分類していくこともありうるだろう。

しかし、個人（あるいは集団であっても）膨大な量に及ぶ文献を読みこむのには、時間がかかる極めて骨の折れる作業であり、一人のアーティストの研究だけに没頭し生涯を費やすことになるかもしれない。

これは「書籍を読む」ということを前提としているからであり、つまり文章の意味を理解しなければならないという前提がある。

文章の内容理解には、統計的自然言語処理などの発達が必要であり、内容を「理解」する人工知能がなければならないだろう。人工知能は機械学習をして、大量の文献データを読み、研究者はそれを巧みに利用していくというのが書籍ビッグデータの扱い方の要点である。

9. まとめ

Google Books Ngram Viewer は、これまでの書籍へのアプローチを一変させるものであるが、それに呼応して出現したカルチャロミクスのような方法は、今後の書籍のあり方、文字文化、自然言語処理などと結びついてさらに発展していくと思われる。「美術史」という分野でも、従来にはない理解の仕方が期待できるだろう。

さらに「Cultural Evolution」(文化進化)を捉える研究が出現してきており、これは進化生物学の手法を用いて、文化を生物の進化のように捉えていくものである。

アレックス・メスーディ (Alex Mesoudi) は著書 "Cultural Evolution:

How Darwinian Theory Can Explain Human Culture And Synthesize The Social Sciences"[註10]の序文でダーウィンの「種の起源」は2009年で出版150周年を迎えたことから書き始めているのだが、彼が自著のなかで展開しているのは、ダーウィンの進化論の応用、つまり「生物学的進化と社会変化」(biological evolution and cultural change)ということに対してである。

また「文化」は様々な定義できるのだが、ここで彼は文化を、模倣や教えや言葉のような社会的な伝達メカニズムを通して他者から得る「情報」であるとしている。

この段階では、まだ文化はまだ広い範囲での文化として原理的に捉えられていて、本稿で取り上げたような200年ほ

どの時間のなかでの美術、それも近現代美術という限定された範囲だけを対象とするものとは異なっている。

だがまた一方で、「文化進化」としてこの200年ほどの美術運動やアーティストの業績を原理的には捉えることもできるし、そこには生存競争のような生物学的な原理もマクロ的に見ると働いている。

そういう意味では、メスーディの視点と本稿では同じ文化現象を、近い思考方法でありながらも、別角度から見ていと言っているのかもしれない。ただこの二極の間にある未開拓の部分は、これからどんどん読み込まれ、解析されていくものになると考えられる。

そして「美術史」の研究もまた従来の方法とは異なる展開を見せるようになるのではないだろうか。

註

- [註1] Fred Brown, "Some Remarks On Impressionism", The Art Journal, 1893, J.S. VIRTUE & Co. LTD, p.103
- [註2] 工藤智行, " 検索エンジンを作る第6回 N-gram と形態素解析との比較",
<http://gihyo.jp/dev/serial/01/make-findspot/0006>
2017/12
- [註3] Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", 2012 The MIT Press, p.601
(注) 同じ例題は Wikipedia 日本語版で「転置インデックス」についての説明に見られる
- [註4] <https://news.stanford.edu/pr/2004/pr-google-121404.html> 2017/12
- [註5] "Quantitative Analysis of Culture Using Millions of Digitized Books", 2011, Science, American Association for the Advancement of Science
- [註6] エレット エイデン, ジャン＝バティースト ミシェル, "カルチャロミクス：文化をビッグデータで計測する" 阪本芳久 (訳)、草思社、2016
- [註7] Sarah Zhang, "The Pitfalls of Using Google Ngram to Study Language", <https://www.wired.com/2015/10/pitfalls-of-studying-language-with-google-ngram/> 2017/12
- [註8] "Culturomics" <https://en.wikipedia.org/wiki/Culturomics> Wikipedia, 2017/12
- [註9] エレット エイデン, ジャン＝バティースト ミシェル, "カルチャロミクス：文化をビッグデータで計測する" 阪本芳久 (訳)、草思社、2016、Kindle の位置 No.1642、Kindle 版
- [註10] Alex Mesoudi, "Cultural Evolution: How Darwinian Theory Can Explain Human Culture and Synthesize the Social Sciences", 2011, University of Chicago Press, pp vii - 1

図版1-7 検索並び出典データ

Google Ngram Viwer, <https://books.google.com/ngrams>
2017/12