

# 「統計的」社会調査法とビッグデータ

——社会学方法論の研究（その5）

水谷史男

抽象化というものに危惧を持ち、データから決して離れまいとする人がいます。ちょうど記述言語学や記述意味論をやるようなものですね。人文科学や自然科学における研究活動を観れば、ごくわずかの例外を除いては、データに依っている度合いが非常に高いことがわかります。人文科学でも自然科学でも、単にデータを整理する類いの研究に留まらず、理論的研究へと進んでいくような知的試みはごく一部分でしかなされていません。アルゴリズムを作っていると安心感があるんだと思います。とんでもない誤りは犯しやうがないですから。

これは思いつきの精神分析みたいなものになるかも知れませんが、こう思うんです。多くの人々は、真に発見できるようなものにはめつたになくて、全ての事柄は、概ね目に映る表面的なものからは程遠いものでもなく、そして、複雑で、もしかしたら驚嘆すべき特性さえ持つようなものなどは存在しないのだ、と信じたがっているのでしょう。そういうものを発見しなくてはならないのだなどとは思いたくない、ということも十分あり得ますね。そして、文の統辞解析用のアルゴリズム等を作ったりしている限り、遠大で抽象的な諸原理や複雑な演繹構造をもった理論、さらに、そのような心的対象を持つ物理的本質は何かという問題などは存在する世界に足を踏み入れずに済むように感じるんでしょう。言語学における行動主義や記述主義が推し進められた原因もこんなところにあるように思うんですが、初めに言ったように、これは総て素人の精神分析に過ぎません。

（ノーム・チョムスキー『生成文法の企て』原著一九八二年、福井直樹・辻子美保子訳、岩波現代文庫二〇一一年、六五〜六六頁）

## 「統計的」社会調査法とビッグデータ

はじめに

- 一 ビッグデータとは何なのか？
- 二 数字とことばへの補助線——チョムスキーの言語論
- 三 情報社会学？——吉田民人の情報論
- 四 社会調査は生き延びるのか？  
おわりに

## はじめに

特定の社会現象を一定の方法で測定して数量的データとし、これを統計的手法で分析するのがこれまで行なわれてきた数量的社会調査だと考えると、そこに大量現象を把握する方法として単純に対象を片っ端から度数を数えていく記述統計学 (descriptive statistics) に代わって、二〇世紀前半、R・A・フィッシャーが創設した統計的推測理論が、社会学においても有力な技法として定着したのは日本では第二次大戦後だったといえよう。今日にいたるまで、われわれはこの推測統計を社会調査の標準教科として学生に教えてきた。

いわゆる標本調査は数十万、数百万、ときには一国単位の国民有権者を大量母集団として、そこから無作為抽出によって選ばれた標本を調査することで、誤差やノイズを管理できる、つまり、誤差を正確に想定された確率分布として計算できるように設計すればよい、という考え方に基づいていた。これができていれば、標本の示す結果から母集団の特性値を推定することが可能である。標本誤差をどの範囲まで許容するかは、標本分布につい

での前提があてはまるなら誤差は確率的に計算でき、たとえば5%なり1%程度に収まれば母集団での特性にもほぼあてはまると判断してよい、というのが教科書の記述である。

それは、ある仮定に立っている。つまり、現実の観測値の背後に、安定的な「無限母集団」が存在すると想定でき、観測値はそこからランダムに選ばれたものと見なす仮定である。この仮定が満たされるなら大数法則が成立する。だが、工場での大量生産品の検査のような場合はともかく、多様な人間が含まれる社会集団を「無限母集団」とみなしてランダムな観測値が出てくるとは考えにくいから、現実に社会調査を行う場合は、ランダム性を確保するために、確率的に誤差を計算できる程度の無作為抽出（ランダム・サンプリング）の技法を駆使する必要がある、というのも初歩的教科書に必ず書いてある。

統計的推測理論の基本的枠組みは一九六〇年ごろまでに完成されたが、折から先進各国で工業化、消費社会化、大衆社会化が進行し、その変動を捕えるためには、推測統計の技法が最も有効だとみられた。工業、農業等における大量生産の場での品質管理と、官庁統計を中心とした調査で、この標本調査の技法は活用された。確率メカニズムを持ち込むことによって、大規模な集団の平均的性質、あるいは構成比を把握できるからだ。そして社会学が二〇世紀後半に「実証的研究」と称してせっせとやってきた社会調査の多くは、やはりこの標本調査を使ったものだった。

しかし、二一世紀の現在、われわれが生きているのはポスト工業社会、高度化し変質した大衆消費社会である。この目の前の社会を正確に捉えるための方法は、今までの標本調査、あるいは推測統計を基本とした数量的調査でよいのだろうか？

統計学者竹内啓は、ビッグデータにふれた文章でこのように書いている。

フィッシャー・ネイマンの統計的推測理論は、二〇世紀の大量生産、大量消費、大衆社会のM A S S論理の支配する時代に最もよく適合したものであったのである。

しかし二〇世紀の最終四半期になって、IT技術が発展し普及すると、規格化された大量生産の時代から、個性的な多種少量生産の時代になり、不良率を抑えることなく、不良品を出さないことが目的とされるようになった。それとともに古典的な統計的推測の方法が必要とされる分野は少なくなった。勿論それが有効に用いられる場合はまだ多くあるが、コンピュータの発達とともに発展した計算科学、情報科学は、統計学以外に数量的情報を処理する多くの方法を生み出したのである。

ところが最近になって、ビッグデータが重視されるようになって、統計学がまたもはやされるようになった。永年統計学の研究に携わってきた者としては喜ぶべきことかもしれないが、統計学を単にビッグデータを扱うための「道具箱」として、使いやすい道具を適当に使えばよいと考えられるのはよくないと思う。或いは道具にしても大工道具ならば、非常に多くのものがあるので、使われる材料や何を作るかに応じて、適当な道具を適切に用いなければならない。それを判断するには、大工としての「腕」が必要である。統計的方法を適切に用いるのには、統計的方法の性質をよく理解し、データ分析の目的に応じた方法を選び、その結果を正しく解釈しなければならない。その判断をコンピューターソフトに任せてしまうことはできない。<sup>(1)</sup>

AI（人工知能）とビッグデータ、これにからむIoT、ディープラーニング、シンギュラリティなどという用語とその応用例は、数年前から企業やジャーナリズムだけでなく政府の関心も呼ぶホットな話題になっている。その中には多分に見当違いの過剰な期待や、万能のマジックであるかのような怪しげなお話も混じって賑わっている。いまのところこれは、新しい技術が未来を拓く、といったような「明るい話題」として語られることが多いのだが、果たして手放しで喜ぶようなことなのだろうか。

楽観的なテクノロジー礼賛に批判的な論者もいて、その中には一九世紀の十年代にイングランドで起きた労働者や農民の機械破壊運動「ラッター」を想起し、彼らは後世まで技術革新を受け容れない愚かな人々として見られてしまったが、技術が社会を変える変動期には光と影があり、産業革命には犠牲を伴ったことを忘れない方がよいと警告する<sup>(2)</sup>。

筆者は人工知能や情報学の専門知識は持っていないし、AIやビッグデータ・ブームに積極的な何かをコメントするつもりも資格もないので、ここではただ、社会学のやってきた社会調査という方法との関わりで、この人工知能ブーム周辺の動向について考えてみることにする。つまり、端的に言えば、この新しいテクニクは、果たして社会調査になにか利用できるものなのか？あるいはもしかしたら社会調査は、もう古臭い使えない技法になってこっちのビッグデータ的方法にとって代わられるというような可能性はあるのか？という問いを立ててみる。

## 一 ビッグデータとは何なのか？

とりあえず初歩的に、「ビッグデータとは何か？」から始める。

大手メディアに流れたごく最近の話題に、囲碁の世界トップとされる韓国人棋士にグーグルが開発したAI (Artificial Intelligence 人工知能) が四勝一敗で勝った、というニュースがあった。手順の組み合わせが十の二〇乗のチェスや、二二〇乗の将棋ではすでに人間の名人にAIが勝っている。だが、囲碁の盤面は縦横一九×一九あつて一手目に取りうる手は三六一通りあり、二手目以降の手の可能性は三六一の階乗になるから、ものすごく膨大でコンピュータで計算させても時間がかかつて、実際上はまだ無理だといわれていた。

盤面ゲームのような例外なき比較的限られたルールであつても、人間の知能に匹敵する正解を導くのは難しいとされていた。しかし、囲碁でもAIが人間に勝てたのは、コンピュータの計算能力が予想以上に高まったこともあるが、それだけではなく、とりうる手とその手の勝負上の評価をコンピュータにすべて計算させるのではなく、別の方法をとったことにある。それは、人間の脳が行っている仕組み（囲碁名人が行う天才的記憶とひらめきの実績「棋譜」のパターンを模倣するというディープラーニングの手法であり、それを使ったことよつて、グーグルのAIは囲碁名人に勝つたと報じられた。

というわけでいよいよAIは人間の知的活動を超えるか、超えないまでもそれと同等の仕事をするようになる、と話は拡大し、ロボットにAIを搭載して合理的で正確な判断をするAIロボットはそう遠くない未来に、頭の悪い人間に代わつて労働現場で中核的意思決定に関与したり、あるいは自動運転自動車がまもなく実用化される

ように、これまでにない明るい未来がAIによって人類に約束されるのだ、というテクノロジーの楽園を語る人があちこちに現われている。

こういう言説に共通なのは、人間の環境への認知能力は身体の疲労や意識の指向性に左右されて、気まぐれや曖昧さ、あるいは誤認・錯覚が忍び込むのに対して、機械であるセンサーは補足に漏れがなく故障さえしなければ、常時観測し記録することができる、だからそれをもとに合理的な判断がAIによってなされれば、その方が人間が行う判断よりずっと正確で的確である、という信念である。そしていま人間を超える能力を持つAIがハイスピードで実現しつつあるということになる。

確かに人間の行う判断にはしばしば間違いがある。単純な疲労や老化からくるミスもあれば、熟慮の果ての失敗もある。それに比べればAIは過去のあらゆる事例や理論的整合性をチェックして出てくるものだから、ずっと信頼できるはずだと。自動運転自動車はほんとうに事故は起こさないのか？という危惧に対して、仮に事故が起きたとしても人間が運転した場合の事故率よりも、自動運転車の事故確率はずっと低い、と説明されるだろう。さて、われわれはこれに納得できるか？

またこれも最近のニュースでとりあげられたもうひとつの話題をあげてみる。

日立製作所が今年（二〇一六年）六月、労働者の幸福感を向上させるAI技術を実用化したと報じられた。これは職場で個人に名札型のウェアラブルセンサーを常時携帯してもらい、各自の行動データを蓄積して幸福感を高めるアドヴァイスをするというサービスだという。センサーには「誰といつ会っているか」という対面情報を記録する赤外線センサーと、オフィスでのデスクワークなど身体動作を記録する可読度センサーが搭載されている

る。人と会っている時なら「話し手」か「聞き手」かまでも記録される。日々膨大なデータになるが、これを日と呼ぶAIで分析し、その結果をスマートフォンを通じて自動でアドヴァイスを行うという。

すでに一三社で実証実験をすませ、システム導入が始まっているという。実は今までは経営管理層の労務管理支援のために、この種のデータをAIで分析していたのだが、今度は従業員自身が自分の幸福度に関するデータを見ることができるといふわけである。たとえば、従業員の平均幸福度が高い日は、低い日に比べ受注率が三四%高いという実験結果が出たという。

開発者の説明によれば、これまでに長期にわたってムードが悪くなったり、逆に充実感を感じていたりする際に、人間がどのような反応をするかを大量に分析するために、一〇の組織、四六八人に幸福感に関する二〇項目のアンケートを行い、その結果を組織ごとに平均化し、その組織が幸せかどうかを数値化する。そして四六八人にウェアラブルセンサーを長期間装着してデータをとったところ、行動の多様性が強いほど、組織での幸福感が高いことが分かったという。

当社の日は、データから自動で学習し、賢くなる。特徴は三つあり、一つ目は、アウトカム（目的）と入出力を人間が定義することである。今回の場合、アウトカムは「一人一人の幸福感の向上」である。アウトカムを基に、関連データを入れていくことで、日が自動で学習していく。二つ目は、人の仮説や問題特有のロジックは入力不要なこと。三つ目は、既存システムに追加し、動作できることだ。

当社は、大量の複合指標の生成と、その中から少ない重要な指標を自動で絞り込む処理を行う「跳躍学習」



技術を開発した。跳躍学習は、強化学習の分野になる。しかし、現行の強化学習は、結果のデータが少ないことに対応できていない。また、特定のニーズに特化してプログラムを開発しているため、汎用的でない。Hは、非常に汎用的に作られているため、一四分野五七案件で活用されているが、全て同じプログラムを活用している。これにより、機械学習やディープラーニングにおいて必要だった教師となるデータ、報酬ロジックなどが不要となっている<sup>(3)</sup>。

この手法が従来の社会調査と違うところは、なんだろうか？

一〇の組織の四六八人に幸福感に関する二〇項目のアンケートをするのは、どのようなサンプリングをしたのか？が詳らかではないが、手法としては社会調査の方法と変わらない。問題はその次に、その四六八人全員にウェアラブルセンサーを装着して、本人の日常で誰といつ会っているか、そこで話し手か聞き手かまでデータにすることが実現していることだ。つまりこれがいわゆる「ビッグデータ」の特徴のひとつである。

従来の社会調査では、あらかじめ設定された質問項目に答えた数値のみがデータであった。それは調査する側が絞り込んだ現実のある側面（変数化した測定値）に限ったデータであるが、それ以外の情報は得られないし初めから捨象されている。われわれはあらゆる無数の情報をくまなく知ることができないし、知ることができたとしても大量過ぎてとても「処理解析」できない。ところが高速処理するコンピュータと工夫されたAIソフトを使えばそれは「できる頭脳」になるのかもしれない。というより、センサーを任意の対象に設置できさえすれば、日々データは自動的に生産され蓄積される。それは特定の目的や研究意図に制約される必要はない。むしろ目的

など設定せずに人の行動や言動を片っ端から記録してしまうことで、それは巨大な記録の集積、つまり「ビッグデータ」になる。

単純に考えても、一人の人間の購買行動の記録、GPSの位置情報の記録、さらにはネットSNSやスマホの通信記録が捕捉されているのなら、個人の住所氏名といった個人情報よりもある意味で物凄いプライバシー情報であるにもかかわらず、すでにどこかにしつかり記録されている。それが使われ解析されているかどうかは、当人には知ることができない。そこから誰かがプロファイリングしたデータ分析結果が、悪用される可能性はないとはいえない。ではいっそ、二〇世紀的なプライバシーシー保護などという時代遅れの配慮はやめて、社会学はこれをデータとして活用できる可能性はないのか？

もうひとつビッグデータとAIがらみの話題をあげてみる。

今年文部科学省が発足させた「第5期科学技術基本計画」なるものがある。科学技術基本計画は、科学技術基本法の定めにより政府が定める五年単位の計画であるが、この第五期計画は二〇一六年一月に閣議決定された。その計画のタイトルは「IoT／ビッグデータ（BD）／人工知能（AI）がもたらす『超スマート社会』への挑戦」。我が国が世界のフロントランナーであるために」となっていて、中心的アイデアはこのBDとAIという新しい技術を、「我が国が世界的な経済競争に打ち勝つ観点から」期待を込めて「超スマート社会」を国として取り組む「society 5.0」と定義する。白書として訴求力のあるキャッチフレーズを文部官僚が工夫したのだから、中身ははなはだ曖昧である。

「超スマート社会」とは何か？白書の説明はこうなっている。

超スマート社会とは、第5期基本計画によれば、「必要なもの・サービスを、必要な人に、必要な時に、必要なだけ提供し、社会の様々なニーズにきめ細かに対応でき、あらゆる人が質の高いサービスを受けられ、年齢、性別、地域、言語といった様々な違いを乗り越え、生き活きと快適に暮らすことのできる社会」と定義している。なお、Society 5.0は、「狩猟社会、農耕社会、工業社会、情報社会に続くような新たな社会を生み出す変革を科学技術イノベーションが先導していく、という意味を持つ」としている。Society 5.0は、具体的には、超スマート社会の実現に向けて、「超スマート社会サービスパラットフォーム」を形成し、複数の異なるシステムを連携協調させることで新たな価値創出を図るとともに、データフォーマット等の標準化や、必要となる人材育成等を実施することとしている。さらに、「超スマート社会」における我が国の競争力強化に向けての知的財産化や国際標準化、基盤技術の戦略的強化としてサイバーセキュリティ技術、IoTシステム構築技術、ビッグデータ解析技術、人工知能技術等の強化を図ることとしている<sup>4)</sup>。

この説明の中にはないが、国民向けの「超スマート社会」の便利さと快適さの説明には、ロボットも頻繁に登場している。以前にもどこかで聞いたような話だと思ったら、二一世紀が始まった頃にひとしきりIT技術の日常生活改善として騒がれた「ユビキタス・ホーム」の夢であった。帰宅する時間に合わせて電気がついたり風呂が沸いたりする「いつでも、どこでも、誰でも、何でもつながるユビキタス社会」を推進すると言ったのは日本の総務省だった。あれはまだスマホが浸透する前だったが、コンピュータを駆使してどんどん生活は便利快適になる、という能天気な楽観的技術信仰に果たして一般の人々がどこまで共感したかはわからない。だが今、「ユ

ビギタス社会」なども誰も話題にしない。

また、二〇〇六年頃に話題になった「Web2.0」の夢も、一億総表現社会というような表現で、誰もがブログで自己表現をし、世界中で何十億人が意見を発表する民主的なネット環境ができると囃し立てた。そこから新しいビジネスチャンスも生まれるという期待は果たしてどうなったのだろうか。ハードとコストの面ではおそらく長足の進歩をしているのだらうと推測するが、移り気なメディア上の議論ははしゃいで終わった感がある。

今度の「超スマート社会」に新しいアイデアがあるとするれば、人工知能（AI）が進化して自動車も飛行機も自動運転するだけでなく、自分の健康管理から余暇娯楽のメニューまでロボットが用意してくれる社会といったイメージなのかもしれない。そして当然その反動に、そんなのいやだ、という意見も出てくるだろう。先にみた一九世紀の「ラッダイト運動」は、愚かなブレモダンの抵抗と笑われたが、もし「超スマート社会」が実際にわれわれの生活の中に浸透したら、「ネオ・ラッダイト運動」も芽を吹く可能性はないとはいえない。「超スマート社会」に抵抗したければ、なにもコンピュータやスマホを破壊する必要はなく、それを使わなければいいだけで、使わない不利益は耐えられないほどのものでもないが、それでもセンサーから逃れるのは電波の及ばない山の中に籠るくらいしかない。それは信念と価値観の問題でもあるが、すでにかなり社会生活に浸透している現実でもある。

たとえば毎日身につけるメガネフレームに、マイクやカメラを埋め込んだセンサーを仕込んでおけば、その人の現在見ている光景や、会っている人物の様子を自動的にネットに送信し、友人に知らせることが可能になる。体温、脈拍、血圧などのデータを常時計測するツールをつけていれば、健康モニタリングも完備する。それらの

データがコンピュータ処理されると、自分のライフ・ログング（行動や生活の履歴）に基づいた的確な診断が送られてくるわけで、高齢者が一人暮らしをする状況が激増する社会には役に立つだろう。

囲碁や幸福度センサーといった話題は、その一端であり、スマホが浸透した現代の情報環境において、なにか人間生活を豊かにする素晴らしいものであるかのような言説と社会的期待が強まっていることは確かだろう。顧客の消費性向に合わせてピンポイントの広告を不断に提供してくれるターゲティング広告や、オンライン・ショッピングの協調フィルタリング（よく似た商品購買傾向を持つ顧客たちの購買歴からお好みの商品をお勧めする機能）など、AI活用の例は今日われわれには日常ありふれたものになっている。人工知能という技術は、コンピュータが誕生した二〇世紀半ばからいろいろと検討されてきた歴史があり、最近急に出てきた話ではない。これに対して、「ビッグデータ」のほうは二〇世紀にはとても考えられなかったものである。

通説に従っておけば、「ビッグデータ」については三つの特徴があると説明される。いわゆる三つのV。つまり、デジタル化による桁違いのデータ量（Volume）がひとつ。これはもう人間が眺めて処理できる限界はとっくに越えている。一メガバイトのフロッピーの時代からみれば嘘のように、USB容量は一〇ギガ、三〇ギガと拡大し、二〇二〇年には、地球上のデータ総量は四〇ゼタバイト（一ゼタは一〇の二一乗、一〇億兆）になるという。とにかく物凄い量のデータが、従来の図書館に積み上げられていた、紙に書かれた情報の数十倍、いや数千倍の勢いで記録されているわけだ。

第二は、データの種類がむやみに多いこと（Variety）で、量だけではなく内容も形式も多様であること。今までの統計データのように限定され整理された数字だけではなく、文章もあれば画像もあり、SNSに書き込ま

れた短文から音声記録、写真・映像や音楽まで、およそあらゆる情報が意図するしないに関わらず、日々記録され利用されている。しかも、ただ多様なデータが別々に並んでいるだけでなく、コンピュータの上で連結し組み合わせ共有したり新たなデータを（その気になれば……）作ったりできることである。

第三は、超大規模なデータ群を取得し上手に処理するためのハードとソフト技術の出現であり、各地で発生する気象データ、交通状況データ、ツイッターの発信データなど、リアルタイムで流れていく膨大なデータを扱えるようになったストリーム・データを処理する速度（Velocity）である。インターネットを前提とした近年のデータ処理サービス、とりわけクラウド・コンピューティングが、ここで解決の道を開いたという。個々のユーザーは、自前のコンピュータにデータやソフトを揃えて組み込んで、あれこれ苦労して難しい処理をするのではなく、専門サービス業者にデータ処理をいわば外注委託するだけでいい。天空のクラウド（雲）の向こうで処理を請け負うサービス業者は、インターネットのどこかで巧みな高速分散処理をおこなう、結果を送ってくれて、ユーザーの要求に応じてくれるようになりつつある。

西垣通の教えるところによれば、人工知能が一般社会でブームになったのは、今回が三回目だという。その第一次は一九五〇～六〇年代で、「正確な思考計算」のための機械としての大型コンピュータで、人間が頭を使うパズルやゲームを論理的

表1 人工知能ブームの歴史

ブーム	キーワード	応用範囲	正確性
第一次 (1950～60年代)	論理	小 (パズル、ゲーム等)	◎
第二次 (1980年代)	知識	中 (エキスパート・システム等)	○
第三次 (2010年代)	統計 (学習)	大 (パターン認識、機械翻訳等)	△

出典：西垣通『ビッグデータと人工知能』中公新書, p.172

に分析するという試みだった。これは頭の体操にコンピュータを使う遊びだった。ただこの段階ではまだ、チェスでもコンピュータは人間の名人に及ばなかった。

次の第二次は一九八〇年代で、これは大量に蓄積された特定の専門的データを、人間がいちいち手で探さなくてもいいように、コンピュータで演繹的に整理し引き出すエキスパート・システムのような試みだった。医療や法律の分野ではこれに大きなメリットが期待された。筆者もあの頃、膨大な判例をコンピュータで即時に検索するシステム構築に情熱を燃やしていた法律家の友人の情熱を思い浮かべる。病気の診断や係争事件の判例をコンピュータが即時に回答してくれるシステムのメリットは、経験の乏しい未熟な専門家よりエキスパート・システムの答えの方が効率的で信頼に値するというユーザーの期待に、確かに応えるものだったと思う。しかしそれは、あくまで医療や法律など特定の体系的な知識と専門家の経験的判断を、コンピュータ上の記憶装置に溜め込んだ大きなファイルキャビネットの高速検索システム以上のもではなかった。

そして第三次が今回の二〇一〇年代になる。前の第一次、第二次と違うのは、扱う大量データの統計処理の導入によって、パターン認識や機械翻訳が可能になり、専門家だけでなく一般大衆に属する人々の具体的な生活に統計的な根拠を与えアドヴァイスまでする、という、実用性にまで一気に飛んでしまう危うさである。しかも、データは単体のコンピュータ記憶措置ではなく、クラウド型ネットワークに日々記憶され、誰が使うかもわからない。人工知能が単に専門家の新しい高度技術であるうちは、無視してもとりあえず問題はないと思っていた社会学は、ここから先に社会学が引き受けるべき課題に、AIがじわじわ踏み込んでくることを予感せざるを得なくなった。

人工知能をめぐるホットなブームの応用を、とりあえず西垣通のあげている成功例とされた試みを参考にみてみたい。

米国ニューヨーク市では毎年、数百もの電力用マンホールが内部で引火して、爆発事故を起こしていた。爆発すれば重い鉄の蓋がものすごい勢いで吹き飛ばされるので、危険きわまりない。だが、同市の電力ケーブルの総延長は地球三周分以上であり、更にマンホールの総数はマンハッタンだけで五万をこえるので、危険箇所の特定は容易ではない。そこで、電力ケーブルの使用年数と、過去に爆発事故をおこしたマンホールの位置情報とを関連づけコンピュータで分析した。そして危険箇所のリストを作ったところ、リストにあげられた多くのマンホールでその後、現実には爆発事故が起きたのである。つまりビッグデータ分析の結果、かなりの確度で、危険なマンホールを予測することに成功したというわけだ<sup>5)</sup>。

これはビッグデータの特徴の第一、とにかく人間の把握の限界を越えたボリュームの大量データをコンピュータは情報処理できるという威力を示した。ここでニューヨーク市がやっていることは、きわめて常識的な手順だが、これまでは人間には膨大過ぎて不可能と考えられた情報処理解析もコンピュータならできるといふ事実である。

近代科学が研究の方法として鋭意確立してきたのは、経験主義、合理主義、実証主義を基本とする実験・観察・分析の手法である。そこで得られたデータの分析を行う意義は、さまざまな現象を説明しそこに働く法則を解明



することにある。研究者は自分の立てた理論仮説をデータによって検証することこそ、知的活動の主流と考えるようになったのである。だが、いま問題のビッグデータ分析処理は、こうした従来の考え方とはかなり異なる特徴があるようだ。西垣通があげているおもな特徴は、次の三点である。

第一は「全件処理」。これは従来の方法、つまり、推測統計理論に基づく標本調査、たくさんのデータから限られた少数のサンプル（標本）を抜きとり、それらを分析してデータ全体の傾向を推し量るという方法とは、基本的に大きく違う。推測統計では、全体（母集団）の縮図を作るように、まんべんなく標本抽出をすることが何より大事だとされる。数値が「正規分布にしたがう」という数学的な性質をもつと仮定すれば、一定数のサンプルの分析によって、母集団の特性をかなり正確に推定することができるからだ。

ところが、ビッグデータはこういったアプローチは不要だと考える。サンプルでなく、ともかくすべてのデータを調べようという。社会調査で言えば標本調査ではなく全数調査。工場の生産管理で言えば、製造機械や部品にセンサーをつけ、計画通りに作られていく全部品のデータを時々刻々、無線で細かく取得し、リアルタイムでコンピュータ処理による集計分析を実行しながら品質管理をする、といったイメージである。

ここで、西垣のあげる「質より量」というビッグデータ分析の第二の特徴が現われる。

センサーは山のように多数設置するので、その一部は壊れたり摩耗して、測定精度が低下している可能性は高い。つまりデータの質は落ちていくのだが、そういうデータは「ノイズ（雑音）」であり、誤差として無視しても、全体的な特性は全件処理によって正確にわかる、というのがビッグデータ分析の基本的な考え方だという。

このことは、製品の品質管理よりも、ツイッターの分析といった例をあげたほうがわかりやすい。ツイッター

で刻々と発生する膨大な発言ひとつひとつの内容を、きちんと論理的に分析するのは至難の業である。だがそれらをとにかく山のように集めれば、多少のブレは相殺され、全体の傾向があぶり出されてくるだろうとみなす。たとえば、ある政治的なテーマについて世論の動向を調べたいとしよう。従来の世論調査ならアンケート項目を準備し、無作為抽出した二〇〇〇人くらいに對し、政策への賛成反対を調査票や電話で尋ねる、といった方法がとられた。だが、ビッグデータ分析では、大量のツイッター発言をまるごと自然言語処理プログラムで分析し、キーワードなどを抽出して、集団的なおよその特性を探りだそうとする。ツイッター発言には俗語表現や文法的誤りも少なくないが、そんな乱雑データはノイズであり、無視してよいというわけだ。つまり、ここには、データの質が多少悪くても大量処理によって正確性が増す、という強い信念があるのである。

かつての全数調査主義、記述統計学が諦めたことを高性能コンピュータを使って復活させたともいえる。しかし、西垣も言うように、「データの質が悪くても、ともかく量をこなせば正確な分析ができる、という考え方は少々乱暴すぎ<sup>6)</sup>」だと思う。データの質は要するに玉石混交であるから、分析結果といっても言語表現データのデジタル変換や処理の仕方次第で、正反対の結果になる可能性があり、信頼度は落ちる。それを克服しようと推測統計学に基づく社会調査では、質問項目や回答処理にじゅうぶんな統計学的注意を払ってデータの信頼性を大事にしてきたのである。

多くの場合、データ分析の目的が、未来の出来事の予測なのはご承知の通りだ。明日の天気だの、火山の噴火だのを予測するといったことである。しかし、当然ながら未来の出来事のデータは存在しない。つまり

過去に蓄積したデータから推測するほかはないのである。したがって予測問題はどうしても推測統計学の領域に入りこんでしまわざるをえない。要するに、ビッグデータの全件処理といっても厳密には、ただサンプル数が増加しただけではないか、という冷めた議論が現われるわけである。

予測の問題はさておいても、超大規模データに対してそもそも全件処理など技術的に可能なのか、という疑問も無いではない。いくらコンピュータが高性能といっても、無限の処理能力を持っているわけではない、コストもかかる。ゆえに、あまりにデータ量がふえれば全件処理は現実味をうしなうのだ。そして、この疑問は、次にのべる第三の特徴と深く関連している。

西垣が挙げているビッグデータ分析の第三の特徴とは、「因果から相関へ」である。

原因と結果の関係にあると想定される事項、つまりあることが原因でしかじかの結果が起きた、という説明方程式をとるのが因果的説明である。喫煙を永年続けた結果、肺がんになった、雨が多かったので水たまりがふえ、蚊が大量発生してマラリアが流行した、といった場合である。Aを原因、Bを結果として両者を因果関係で説明するには、あらかじめどれかを独立変数、別のどれかを従属変数と想定してそこに絞って測定をしてみる必要がある。事象を説明する論理的思考として、因果関係のモデルは代表的なものであるが、原因は一つとは限らず、さらに結果を導く過程で働く条件がいろいろあり得る。ところが、ビッグデータ分析では、そんな面倒な手続きはやめ、コンピュータで全部を一気に因果関係ではなく相関関係を調べればよいという考え方がもてはやされているのだという。

相関関係というのは、単にAとBのデータの統計的な関連性を示すものである。たとえば、身長と体重のようにAが高いものはBも高いという正の相関、あるいは運動量を高めれば皮下脂肪が減るといったAがふえるとBが減るといった負の相関があるかどうか、データから確かめることができる。だが、身長とIQ（知能指数）のように、明確な相関関係がみられないものも少なくない。

この相関関係を利用するだけでも、確かにいろいろな実用に役立つ。だが、ビッグデータ分析の魅力は、常識を超えた相関関係を発見することで、有効な行動をとれるということなのだ。たとえば米国でグーグルが行った成功例として、「咳止め薬」や「解熱剤」といったキーワードによるウェブ検索頻度を調べて、どんな地域でインフルエンザが流行しているかを特定できたという。検索キーワードの使用頻度とインフルエンザ感染の時間的・空間的な広がりとの相関関係を分析して、米国の公衆衛生当局よりも早く、インフルエンザ流行の予測に成功したのだという。

検索キーワードとインフルエンザとの相関関係をとってみようというのは、コンピュータの腕力でこそ可能な独創的発想である。ここから、手あたり次第にさまざまなデータのあいだの相関をとってみれば、必ず何か新たな発見につながる、という発想が出てくる。しかし、二者の相関だけではたまたまそうなつたにすぎず、データの種類や件数を増やすと組み合わせ数は飛躍的に巨大になる。検索キーワードとインフルエンザ流行の相関分析にしても、グーグルの研究チームは四億五〇〇〇万にもほる膨大な数式モデルを使って分析作業を行ったという。このようにおおよそ、データの相互関係を二変数で分析するだけでもたいへんなのに、もし、関連の深い四人数ずつのグループを見出そうとして分析作業をすれば、対象者が一人でもその組み合わせ数は約四〇〇兆通りに

およぶことになる。将棋のような限られたルールのもとに一手ごとの可能な展開を計算することはできるとして、異種データ相互の相関関係を分析するには、複雑多岐な組み合わせを、あらかじめ整序することなしに片端からシラミ潰しに計算していくなら、そのコンピュータ計算は想像をはるかに超えた量になるはずだ。

さらに、「因果より相関」というアプローチには、計算量だけでなく、いつそう本質的な問題点がひそんでいると西垣は言う。およそ、「相関関係さえ見つかればいいんだ、結果がわかれば理由なんていらぬ」という乱暴な手法は、単純で反知性的なおいがする。因果関係をたどるのは面倒くさいし、非効率かもしれない。だが、いかに苦しくてもそれは、「思考する存在」である人間にとって、かけがえのないものであり、もし、コンピュータがわれわれの直観や常識に反する相関関係をはじき出したなら、それを安易に信じて行動することは適切なのか。むしろ疑って立ち止まるべきではないのだろうかと述べている。<sup>8)</sup>

テクノロジーで問題を片付けようとする工学的思考には、なから違和感を持つ社会学者としては、そもそも「因果より相関」というアプローチでは、なぜそうなるのか、また問題そのものが孕む価値や目標について、思考の対象から外してしまうものに見える。

また、西垣はデータから「帰納」する論理に対して、ビッグデータ分析では「仮説推量 abduction」という処理をするが、それには誤りが含まれる可能性が排除できないと論じている。<sup>9)</sup>

コンピュータが人間のように自ら学習し思考して、正しい判断をするといつても、個々の事実を一般ルールに照らして個別条件を導く仮説推量をしているだけなら、そこに他の条件が働いていることに気づくことができない。これを社会調査で考えてみれば、二変数の相関があつても、それはみせかけの疑似相関である可能性があり、

第三変数、さらにはそこに影響を与えていると考えられている多数の媒介変数を組み入れて考えなければならぬ、ということとは初歩の常識である。多変量解析のさまざまな技法は、そうした考慮の数量分析の基礎にある。

ベルギーの数学者、A・J・ケトレー（一七九六—一八七四）に始まる近代統計学が追求してきた、当時の新しい研究分野である確率論と統計学は最小二乗法などの形で主として天文学に應用されていた、ラプラスは確率論を社会研究にも應用することを考えていたが、ケトレーはこのアイディアに基づき「社会物理学」の名で研究を開始した。彼の目標は、犯罪率、結婚率、自殺率といったものの統計学的な法則を理解し、他の社会的要因の変数から説明することにあつた。このような発想は当時の学者の間に議論を巻き起こしたが、一八世紀以来の「神の秩序を数学的に明らかにする」という思想に対して、個人の行動に基づいて科学的な法則性を追究した点で際立ったものだった。彼はそれを社会物理学と呼んで、「平均人」(*l'homme moyen* : 社会で正規分布の中心に位置し平均的測定値を示す)という概念で説明した。<sup>⑩</sup>

竹内啓はビッグデータに統計的方法を適用するに当たって、四つの段階を経なければならぬと指摘している。それは、一、データの吟味、二、モデルの選択、三、手法の選択と適用、四、結果の解釈と判断、である。とくに第一段階のデータの吟味という点で、ビッグデータには問題がある。

ビッグデータは莫大な量のデータを含んでいるが、量が多いというだけでは、分析の目的に対応した良質な情報がふくまれているとは限らない。ビッグデータは、ふつう何らかの管理業務のために記録されるのであって、統計的情報を得るために作られるものではない。従つてその対象となつている集団は、分析の目的

からすれば偏っているということが少なくない。例えば特定の店のPOSデータは、当然その店の顧客に関するものに限られるから、顧客の性、年齢、所属、職業などによって偏ったものになる。また業務記録はその主要な項目についてはほぼ正確であるが、それとともに記録されている項目については偏りが大きいことがある。所得や資産については過少に申告されることが多い。また一つの対象について記録される多数の項目について記入が欠けている、いわゆる欠測値が多くあることも少なくない。その場合存在するデータだけを単純に集計すると、結論が偏る場合がある。またデータの中には最初から不適切なもの、或いは根本的にまちがっているものがあり得る。それらは本格的な分析の前に除いておかねばならない。データの吟味はデータ解析において絶対必要である。データが膨大である場合には、その一部を標本として取り出して予備的解析を行うことも必要である。<sup>(1)</sup>

また、こうした統計学的な観点とは別に、大黒岳彦が指摘するように、ビッグデータについて単なるビジネスに資する道具としてしか見ようとしないうちの多くの誤解があり、こうした誤解はビッグデータが指し示している社会の基底的次元での構造変動から目を逸らすことになる、という点は重要である。<sup>(2)</sup>

以上、おもに西垣通氏の所説を追って、ビッグデータとそれを処理するAI技術の現状についてみてきた。ビッグデータという言葉のあいまいさにさまざま危惧を感じるものの、AI技術の可能性という点では、もはや現実の応用分野でこれを反対方向に押しとどめることはほぼ不可能だし、技術的に進むところは進んでいくだろうと筆者には思える。おそらく学術研究とは別の場所、とくにビジネスや金融などの分野では放っておいても、AI

技術の利用はスピードを増していくに違いない。そこで、本稿にとっての問題はもっと限定された場で考えてみたい。

## 二 数字とことばへの補助線——チヨムスキーの言語論

コンピュータ技術がもたらしたAIとビッグデータの問題を論じているのに、言語学を持ち出すのは奇妙だと思われるかもしれないが、まずこういう設定をしておきたい。

「データ」という言葉の語源は、datumという、ラテン語の「与える」dateという動詞の過去分詞形「与えられたもの」から来ている。人間が知りたいという事象、それを記録しておいた「与えられたもの」をデータと呼ぶとすると、コンピュータが直接に扱うのは、言葉ではなく数字である。いうまでもなく、数字や単純な記号だけなら、パリだろうがニューヨークだろうが、東京だろうが時間、空間、文化の差を無視して共通の記述に使うことができる。序数奇数を問わず数字は単に量を表示するだけでなく、アルゴリズムを与えて演算処理させると、最適解をはじめ出す。コンピュータの出發は二進法を応用して電子回路で計算させる機械から始まったことは知られている。ということとは、コンピュータが得意なのは、「与えられたもの」としての数字で計算することで、人間が使うことばのうち数字以外の言語、文字で書かれる文などをそのままでは扱えない。通常ある場所で行われていることばは文法や語彙がそれぞれ違っていて、それが意味するところは数字のように一義的ではなく、表現に無数ともいえるヴァリエーションがあるから、その違いを認識するのはコンピュータにはひどく厄介な作業



になる。

そこで、ことばで認識し記録したデータを人間が数字に置き換える、数量化の作業が必要になる。たとえば社会調査がいつもやっている回答をカテゴリーごとの選択肢に落とし、それに数字をふって一定の尺度にしたがつて数量化するやり方がある。あるいはことばで記録された事実や発言の文を、用意したコードにふり分けて番号を付ける。それは数字に変換されているからコンピュータに読み込ませることができると言える。あとは計算の速度だけの問題である。しかし、この数量化の作業はふつう人間がやらなければならないし、そこには言語に関する恣意的で文化的、あるいは微妙な「文学的」解釈が入り込むことは避けられない。できれば、そういう個別言語のもつ差異は消去しておきたいとIT情報学者は考えるのだろう。

そこで、AIとビッグデータの議論では、もっとシンプルに言語情報を数量化するのではなく、センサーでデータを採る段階で会話も映像もすでに数値になっており、自動的にデジタル情報や機械翻訳で計算可能な形にしておけばよいと考える。あとは計算能力・速度とアルゴリズムをどう工夫するかの問題であるとする。技術の指数関数的進歩を謳う「ムーアの法則」よろしく、それはすでに人間の手にするところとなっているのだ、と。

しかし、これは技術の進歩がつねに人間の幸福利益にプラスに働くという根拠の希薄な妄想でないとは言いきれないだろう。そのことを考えるために、もう一度言葉の問題、言語学を覗いてみる。人工知能とビッグデータの社会への影響というものを、冷静に考えようとするとき、この問題に当初から示唆を与えたチョムスキーの言語論が、補助線になるかもしれないと筆者は考えている。

ユダヤ系アメリカ人、ノーム・チョムスキーは二〇世紀言語学に革命を起こした人として著名であり、一般に

はヴェトナム戦争を推進したアメリカ政府に対して批判的なりベラル派知識人として、ジャーナリズムでもすでに世界に知られた名前である。でも、その言語論、変形生成文法と呼ばれる理論は、いかなる由来、いかなる学問的意図のもとに形成され、革命と呼ぶに値するものなのか、われわれは十分に理解しているとは到底言えない。

近代科学というものの見方・考え方が、広く体系的理論的に整備されたのは、一九世紀なかばの西欧で、自然科学の発見・成果が疑いもなく現実の変化として人の生活を変えることが目に見えるようになった産業社会以後である。電気、動力機械、鉄道、自動車、飛行機、軍艦、大砲。それらを生み出したのは、実験と観察、現象を説明する理論とそれを応用する技術の結合だった。その方法は、物理・化学から生物学・地学・医学におよび、さらに社会現象にまでおよんで次なる二〇世紀に発展する。しかし、人間が話し、書き、読む言語の研究は、どんなに精密に研究し、世界中の言語を採集して記録してもまっとうな社会科学にも入れてもらえなかった。経験科学として人のことばを研究するとしたら、まずは世界中で話されていることばを収集記録し、そこからなにか法則や理論を導き出す以外にないと考えたのは当然とも言える。チョムスキーの言語理論が、革命とも呼ばれたのは、この前提を覆したからだという。

科学研究における経験主義の帰納的方法と、論理主義の演繹的方法との対比は、さまざまな領域分野において問題の焦点になると思うが、言語学の場合、世界の言語の採集を丹念に蓄積することから辞書の作成や音韻論、文法規則の系譜論などが経験的に行われてきたという点では、帰納的方法が主流を形成したのだろうと思う。これに對して、普遍主義・論理主義の言語論は西欧ではつねに存在したが、それが二〇世紀にはつきりと表明されたのはチョムスキーになる、といっている。たとえば、『統辞構造論』第三章第一節で、チョムスキーはこのよ

うに書く。

言語に関するこの考え方（有限状態文法）は、極めて強力で一般的なものである。この考え方を採用できれば、話者というのは本質的にここで考えている機械であると見なすことが出来る。文を産み出すとき、話者は初期状態から始め、文の最初の語を産み出し、そのことによつて2番目の語の選択を制限する第2の状態に移るといふような形で進んで行く。話者が通る各々の状態は、発話におけるその地点での、次の語の選択を制限する文法的制約を表している。

言語に関するこの考え方が持つ一般性とコミュニケーション理論など関連分野における有用性を考慮すれば、英語のような言語、あるいは数学の形式的された体系の統辞的研究において、この観点を採用することによつてもたらされる帰結を探求することは重要である。そして、英語に対して有限状態文法を構築する試みは、たとえそれがいかなるものであつても、最初からすぐに重大な難点や厄介な問題に直面することになるが、このことは読者にも容易に確かめてもらえるだろう。但し、英語に関する次のようなより一般的な言明を考慮すれば、今述べたことを実例によつて示す必要もないであらう。<sup>(13)</sup>

ここでチョムスキーの言う「有限状態言語 (finite state language)」とは、有限数の異なる内部状態 (internal states) のうち、「ある機械」がいずれか一つを取ることができると仮定したとき、産み出された言語のことを指す。そして、文を産み出すその機械を有限状態文法 (finite state grammar) と呼ぶ。ある言語、それが産み出すこ

とのできる文の集合を定義するものが、有限状態文法になる。なぜ文法が必要か、といえば、語の列である文は無限にたくさん存在するのに対し、文法は必ず有限であることが求められるからである。

チョムスキーが拒否する記述言語学は、発音であれ、文法であれ、実際に話されている言語の経験的研究によって得られたデータから帰納したものを基礎にする。世界にはさまざまな異なる言語が併存しており、言語の研究はそこから出発すると考える。しかし、チョムスキーはそうした努力は必要ないと主張した。話され使われていることば（コーパス）の採集は、言語の研究にとって些末で末梢的なものである。コーパスとは、ことばの能力を用い、その力を行使（perform）して、その実演（performance）として現われた、過程の末端に現われる結果にしからずぎないのだから、それをいくら集めても、部分的なものでしかない。むしろ、追及すべきはより本源的なコーパスを産み出すそのもとにある、言語を使う人間の能力（competence）に注目すればよい、と考える。

そこで登場するのが、この能力がそなえている基本的な（言語）形式があらゆる実際の言語表現——つまり表層構造をつくり出しているもの、「深層構造」になる。それぞれの言語は、この深層構造を、一定の規則にもとづいて変形し、表層構造において、いわゆる言語として実現する。この深層構造と表層構造との関係、とりわけ深層構造を表層構造にもって行くのに、どのような変形規則が使われるか、それを明らかにするのが文法の仕事だと考えた。これは言語研究における経験主義・実証主義の否定である。

この深層構造というのは、いくつかの命題（proposition）から成るひとまとまりであって、それにはいろいろな組み合わせがある。その命題の基本的な形は単純な「主語＋述語」であり、それ以上には分解できない文の骨格のようなものと考えられる。チョムスキーが説明のためにあげた実例は、たとえば次のようなものである。

Dieu invisible a créé le monde visible. (仏語) / God created the visible world. (英語)

目に見えない神が、見えるこの世を作った。

表層構造としては単一の文のかたちをとっており、通常の文法ではこの文は、invisible God (目に見えない神) という一つの主語と created the visible world (見えるこの世を作った) という一つの述語から成るとされる。しかしチョムスキーは、オトになって出てくる表層ではそうになっているが、深層におけるそれ以前のものかたちでは、「目に見えない神」がすでに、「主語＋述語」の観念をそなえていると解釈する。このような考え方にしたらえば、表層構造において、見かけは一つのこの文が、深層構造では次のように三つの命題に還元される。

一' Dieu est invisible. (God is invisible.) 神は目に見えない。

二' Il a créé le monde. (He created the world.) かれはこの世を作った。

三' Le monde est visible. (The world is visible.) この世は見える。

こうして描き出されたそれぞれの文は、もうそれ以上分解することのできない基本命題だという。

これらの命題は、人間がもの言うとき、そのままじかには現われないが、「精神に現前している複合観念のうちに入っている」<sup>(14)</sup>。そのことは別の個所では、「深層構造は含蓄的 (implicite) であって、それは表現されることはなく、ただ精神のなかに表示される」。深層構造は口に出されることはなく、ただ心のなかでなんとなく思われるだけにとどまる。それは「純粹に心的 (purely mental)」な過程なのだから。<sup>(15)</sup> この「心中に思われているもの」が「一定の精神的操作」によって表層構造として姿を現わす。その深層が表層へと変形する過程をいくつか

の規則として引き出し、それをチョムスキーは「文法」と呼ぶ。つまり、文法とはチョムスキーによれば、ことばそれ自体の中から現われ、それ自体の中に見出されるのではなく、心（それも脳の働き）とことばとをつなぐつなげ方の方式を記述するものだということになる。それではチョムスキーにとつて、深層構造と表層構造のいずれが言語の本質、本物であるのかといえば、深層構造の方である。

構造ということばを使っているが、これは構造主義言語学を唱えたソシュールの立場に根底から批判を浴びせるものとなっている。田中克彦はそれをこのように言う。

ある文の深層構造とは、ある文の底にひそんでいて、その文の意味を決める抽象的な構造 (structure) である。表層構造とは諸単位から成る表層の組織 (organization) であつて、これがどういふオートとなつて出てくるかを決め、実際の発話の形式、その受容され、あるいは意図された形式に従わせる。ここでは、深層構造は構造であるが、表層構造は諸単位からなる組織（＝組み合わせ）であつて、それは構造の名にあたししない。すなわち、深層構造こそがまともに相手にできる対象だと述べているのである。

我々はこのことばを聞いたときの構造主義者の驚愕がいかばかりであつたかを想像しよう。ひたすら眼前に置かれた、現実の言語のかたちを目をこらしていた、そのような節度ある態度こそが「まさしく、実際の言語運用に関する理論の発展を妨げている」と言うのだから<sup>(16)</sup>。

科学研究における経験主義の帰納的方法と、論理主義の演繹的方法との対比は、さまざまな領域分野において

問題の焦点になると思うが、言語学の場合、世界の言語の採集を丹念に蓄積することから辞書の作成や音韻論、文法規則の系譜論などが経験的に行われたという点では、帰納的方法が主流を形成したのだろうと思う。これに對して、普遍主義・論理主義の言語論は西欧ではつねに存在したが、それがはつきりと理論として表明されたのはチョムスキーになる、といっている。

言語学では、通常その分野を意味論 (semantics)、語用論 (pragmatics)、そして統語論 (syntax) に分け、さらに派的に音声学 (phonetics) や語彙論、文字論、比較言語学などが展開するのだが、チョムスキーはもっぱら統語論 (構文論とも訳される) 中心の生成文法を開拓したといわれる。意味論が、語・句・文・テキストなどの記号列が表す意味について論じるのに対して、語用論は、実際の発話や文脈に依存した記号の使用を問題にし、統語論は、その記号列の構成について論じるものとされていた。

たとえば、Colorless green ideas sleep furiously. (無色の緑色のアイディアは猛り狂って眠る) という文は、語の文法上の位置は正しいから統語論的には問題ないが、意味論からすればめちゃくちゃで意味をなさないことになる。この違いは、「緑色のアイディア」のような表現が、意味論的にはわれわれが日常的に経験している事象の中には見つからないことを知っているからだ。

「近代言語学の父」と称される F・ソシュール (一八五七―一九一三) は、社会的文脈から言語というものを切り離して研究する方法を樹立した人といってもいいが、そのためにはことばを話す個人と社会をどのようなものと設定するかが鍵になる。ソシュールは、デュルケームが始めたフランス社会学の成果のなかにある、個人を超えた「社会的事実」という概念に啓示を受け、自分の理論に言語共同体 (la communauté linguistique) と

うものを持ち込んで、ことば一般のなかで「社会的に制約された部分」、「その社会のメンバー相互の間に交わされた一種の契約にもとづいてはじめて存在する」ラングという概念を確立した。<sup>①</sup>

ソシユールは、言語学とはこのラングを、「社会的生活の内奥において研究する」ものであるとしたが、言語学がその対象を限定し、科学になるためには、言語の超個人性、個人にとって外的であり、強制的であるというこの性質が必須であった。だが、ある言語が存在するためには、話す大衆 (*masse parlante*) が必要である。ソシユールは、社会的事実としての言語に基礎をあたえるために言語共同体を設定したが、その半世紀後に、チョムスキーは言語のあり場所を、社会や大衆などではなく、生物的個体の中に閉じ込めた。

ソシユールではラングとパロール、概念と音響映像の対を言語がむすびつけ構造化すると考えるのに対し、チョムスキーではそれが心の中で、つまり脳の機能として可能なかを問題にする。だから、それは社会の中で言語の体系がどう作られるかではなく、ことばを話す人間の能力を構文や文法の生成として問うシンタクスの問題になる。そこには、人間が生まれながらに持つ言語の能力が、新しい文を自由に作り出す「創造性」をもつという思想が「デカルト流」の言語観として理論化される。

言語に対するチョムスキーのこのようなアプローチは、人文科学あるいは社会科学に分類されていた言語学を、基本的に自然科学に近づけていく。「人が言葉を話す能力をもつのはなぜか」というチョムスキーの問いは、文化や歴史の個別性ではなく、時空を超越した普遍性に指向するところから、記号と論理をあやつる人間の思考能力の問題になり、生物学的基盤につながるからである。つまり、すべての言語が共通に持つ普遍的な一般文法は、論理の普遍性によって根拠を与えられ、その論理はまたさらに、種としての人間に特有の能力を基礎に持つ。種



とは生物中の、ある特徴を共通にするグループのことを言うわけだから、すべては結局、生物学的な基盤に還元できることになる。

西洋語にはなぜ日本語にはない冠詞というものがあるのか？ しかもフランス語では冠詞が男性名詞、女性名詞などで変化するのか？ など、ある言語特有の規則はいろいろ違いがある。それには合理的な理由はなく、ただそういう規則になっているから、というだけで、その根拠を考えるとその言語を話す人々がいる、つまり「話す大衆」がいるからという社会的事実に着する。ソシユールはその「話す大衆」を仮定しておいて、言語の作り作られる構造を考える。でもチョムスキーは、そういう個々の言語の使われ方とか、差異とかいうものに興味がない。いわば言語をどんだん骨だけに分解してしまつて、変形規則という次元にまで抽出してしまえば、フランス語だろうが英語だろうが、何語だろうがみんな同じというところまで持つていく。

こういうあたりの思考様式は、社会学でも二〇世紀の半ばに「交換理論」というのが出てきて、人間の社会的行動を複雑なものとするのではなく、基本的に単純ないくつかの心理的命題に還元してしまい、あとはそこに働く条件だけで全部説明可能だとする、G・C・ホーマンズのような社会学者の理論が思い出される。

しかし、チョムスキーのように構文の形式だけを、内容から切り離して抽出してくるのは困難な言語もある。たとえば中国の古典語（漢文）は、内容を捨象して文の論理形式（文法＝構文論）を考えるのが、たいへん難しい。中国古典語は、個々の単語の意味解釈に依存しない、むきだしのシンタックスは成立しない。文法は、品詞の別のない、概念だけの単語に依存していると見えるからだ。ギリシャ語とかラテン語のように、極めて発達した形態論をそなえた言語を見なれたヨーロッパ人には、「名詞と動詞の区別すらない」「極度に不完全な言語」で

ある漢文と、その言語によって達成された高度な文明との不可解な対照の前にとまどつたはずである。

なるほど。漢文には形態上、名詞、動詞、形容詞、接続詞などの区別がなく主語と述語もなく、変化しない。音声や会話には無縁で、ただ文字があるだけ。並ぶ順番には規則はあるが、漢字という表意文字が、状況や文脈に対応してなんらかの意味を醸し出す。日本では漢文を日本語に読み下して、いわば文法的変換をやってしまう。これはチョムスキーの変形だともいえないことはない。ために、論語の一節でやってみると……。

為仁由己 而由人乎哉 顔淵曰、請問其目 子曰 非礼勿視、非礼勿聽、非礼勿言、非礼勿動。

（『論語』、顔淵 第十一 顔淵仁を問う）

読み下し…仁を為すは己に由る。而して人に由らんや。顔淵曰く、請ふ、その目を問はん。子曰く、礼に非ざれば視ること勿れ、礼に非ざれば聴くこと勿れ、礼に非ざれば言ふこと勿れ、礼に非ざれば動くこと勿れ。

読み下しの方は、主語と述語、そして仮定や命令など日本語の文として意味の通るように操作し注釈を加えている。レ点や振り仮名のような工夫は日本で加工したもので、中国ではこの文字をそのまま読むわけだ。シンタックスの純粹化に成功した漢文。

個別言語を持つ余計な装飾的形態、接続音など、文法的な小道具の欠如のゆえに、漢文は、より言語外の論理そのものへと読む者の精神を集中するかもしれない。チョムスキーの深層構造は、論理的な関係のみを示すかぎりのシンタクティックな構造であるから、当然形態論を含まない。それは表層への変形の過程ではじめて登場してくる。このように考えれば、チョムスキーの深層構造は、漢文に近いものと言えなくもない。ただしそれは、英語の構文を頭に置いているのだが。

チョムスキーの理論が言語学で革命的なものとされ、その影響は日本にもおよんだ。それは八〇年代のコンピュータ科学に顕著だったと西垣通は述べている。

私事にわたって恐縮だが、門外漢である私がチョムスキーに興味をもったきっかけは、その言語理論がコンピュータ科学と深い関わりを持ったからである。本書が書かれた八〇年代は、人工知能研究の全盛期だった。ヒトの言葉を理解し、翻訳や会話をする第五世代コンピュータの夢が本気で語られていたのである。当時のコンピュータ科学では、主にチョムスキー理論がヒトの言語のモデルとして用いられた。たとえば、チョムスキーの文法にしたがって文章を解析・生成する機械翻訳システムも多く開発されたのである。

コンピュータ科学がチョムスキー理論と親和した理由は、まずその数理的な表現形式のためである。チョムスキー理論における言語モデルは、コンピュータのプログラミング言語に酷似した形式的厳密性を持っており、それゆえコンピュータ科学者には扱いやすかったのだ。だがそれだけではない。「深層構造」という考え方が、野望に燃えたエンジニアを魅了した。ヒトの言語表現は一見さまざまのようだが、深奥に共通の「普遍的な意味」があるとチョムスキー理論（少なくとも初期の理論）は述べる。とすれば、うまくするとロボットのヒトの言葉を理解・表現させることもできるはずだ。つまり、深層構造をコンピュータの中で表現し、入力文を外国語に翻訳したりできるだろう、というわけである。実際、たとえば、R・シャンクスの概念依存的構造というモデルでは、入力文の意味をプリミティブな「概念の型（行為、物理的対象、空間位置、時間位置、行為修飾、対象属性など）」の構造体として表現しようと考えた。述語に対応する「行為」は、さら

「統計的」社会調査法とビッグデータ

に十個あまりの基本行為に分類される。たとえば、「所有関係を移動する (give, buy:...)」とか、「物体に力を加える (throw, shift, kick:...)」などである。

この他にも、八〇年代にはさまざまな大同小異の取り組みがなされたが、成果は期待したほどではなかった。ヒトの言葉を真に理解し表現できるロボットは現在でも存在していない (ちなみに、九〇年代以降のコンピュータ科学では、チョムスキー流の文章構造の解析・生成よりもむしろ、多くの用例を集め、文章パターンの合致によって機械翻訳を行うなど、実用的な「コーパス・ベース」の方向が模索されている)。

私は八〇年代、当時のコンピュータ科学のあまりにナイーブな言語観に強い反撥を覚えた。そこには明らかに行き過ぎた普遍主義・論理主義が見られたからである。<sup>(18)</sup>

チョムスキー二八歳の代表作『統辞構造論』の中で、注目すべきはその中身ではなく、冒頭の「まえがき」末尾にある謝辞である、とクリス・ナイト (英国生まれの人類学者で政治活動家) は言ったという。<sup>(19)</sup> チョムスキーがMITでこの研究をすすめた資金の一部が、アメリカ合衆国陸海空軍によって提供されたことに感謝している。その箇所を引用すると、「この研究 (すなわち『統辞構造論』) は、アメリカ陸軍 (通信部隊)、空軍 (航空研究開発本部科学研究局)、そして海軍 (海軍研究部) からの支援を一部受けている。さらに、米国立科学財団とイーストマン・コダック社の支援も部分的に受けている」。<sup>(20)</sup>

アメリカの軍部が言語学という一見おそ軍事とは無関係に見える研究に資金を出したのは、当時の米軍がそのテクノロジーの要請として、チョムスキーの研究に期待すべき何かがあったわけで、それは端的にコンピュー

タ言語に通じると考えたからだろう。そこには、人間がことばという気まぐれで誤解を生みがちな手段を、もつと一義的で間違いのない記号に変換するものにできるのではないか、という期待がある。軍事上の作戦命令を兵士たちに誤りなく伝達するには、個性を持った日常言語では心もとない。英語もまたさまざまな装飾や言い回しの綾に彩られた言語のひとつであり、研ぎ澄ますべきは文学的な洗練ではなく、伝えたい核文だけである。それは意味論的アプローチではなく、構文論の自然科学的アプローチによってなされる。

そのような期待は、もちろんそのままでは果たされなかったが、二一世紀のAIとビッグデータの進展で再び蘇るのかもしれない。ディープラーニングの戦略は、チョムスキーの言語論がそのままでは使えないところから始まっているようだが、コンピュータ言語の普遍性をより実用化するうえで、チョムスキー理論が機械翻訳や脳科学の進展にきわめて大きな影響を与えたことは確かだろう。二一世紀科学における、ことばと数字という問題で、西垣がかつてチョムスキー理論に感じた、行き過ぎた普遍主義・論理主義に加えて、いまやインターネットが世界を覆う情報社会の新段階には、これはあまりにも「アメリカ的」覇権主義の片棒を担ぐのではないか。

### 三 情報社会学？——吉田民人の情報論

ビッグデータというような少々空想的で壮大な話と、チョムスキー言語論という特殊な世界を、門外漢であることを省みずに一瞥してきたが、ここで少し、日本の社会学という狭く囲われた世界に立ち戻ってみよう。

現代の社会学者で「情報」について本格的に考察を進めた人として、われわれには吉田民人という名がすぐあ

がるであろう。一九九六年に「日本社会情報学会」と称する学会が複数設立され、これとは別に「情報社会学会」も二〇〇五年の四月に設立された。「情報社会学 (Infosociology)」を提唱する公文俊平によれば、先行する二つの学会 Socio-Information Studies (J S I S) 、 Social Informatics (J A I S) は理系色の強い「情報学」に重点を置いていているのに対し、「情報社会学会」は「社会学」に力点があるのだという。つまり「情報社会学」は最広義には「情報社会学の学際的総合的研究」を意味し、最狭義には産業社会での「富のゲーム」に代わって情報社会で新たに普及するようになると思われる「智のゲーム」を研究対象とする社会科学の一部門を意味するとい<sup>(21)</sup>う。

情報社会の社会学という構想は、公文によれば、理論経済学から出発した村上泰亮（一九三一～一九三）と理論社会学から出発した吉田民人（一九三一～二〇〇九）の仕事の延長線上に登場したものと考えられる。コンピュータ情報処理から発展する理系の情報学に対して、村上や吉田の考えた「情報」社会は、かなり壮大な歴史的視野に立っている。

村上も吉田もすでにこの世にないが、社会科学的情報論研究というものを考える上で、ここで吉田民人の情報科学論について、少し振り返っておこう。

吉田は先ずクーン流の、一七世紀に起きた「大文字の科学革命」が、二〇世紀後半にも起きていると見る。一七世紀の科学革命が、「物質」ないし「物質⇨エネルギー」の法則を追求するものであったのに対し、現在進行中の第二次科学革命ではそれが「情報」を中心に展開すると考える。この場合の「情報」は、ヒト・ゲノムのような遺伝情報から、脳神経系の生命活動、人間の行う秩序認識や思考における物理自然科学とは別種の情報、言

語的情報までを含むもので、自然科学の秩序形成原理が一元的な「法則」と呼ぶものであるのに対して、独自の論理・数学的構造をもつ「プログラム」を対置する。一種の飛躍、キーワードとしての情報を、科学革命を導くものととらえる。

この第二次科学革命は三つの柱からなっているという。

- (1) 「設計科学 designing science」の提唱
  - (2) 「情報」範疇の追加
  - (3) 「プログラム」原理の導入
- がそれである。

第一の柱としての「設計科学」というのは、従来の近代科学が「自然の認識」を目的として、あくまで自然現象の記述・説明だけに禁欲的に自己限定していたのに対し、「ありたい姿」や「あるべき姿」を設計するという「自然の設計」にまで踏み込むことを科学の目的にする。

第二の柱、「情報」範疇の追加というのは、自然を構成している「物質Ⅱエネルギー」に対して、「非記号情報および記号情報」を追加するということで、吉田の表現では、「物質Ⅱエネルギー」はアリストテレス哲学の「質料」にあたり、「非記号情報および記号情報」はアリストテレスのいう「形相」を科学化することになる。非記号情報の定義で、「差異／パターン」というのは、ソシユールの差異・差異化をとりいれ、「パターン」は計算機科学の「ボタン認識」から拡張して、この「差異とパターン」が同時成立すると考える。「記号情報」は、DNA情報を原型として言語情報やデジタル情報にまで進化するという。これができるのは生物相と人間に固有のものである。

る。DNA情報は「AGCT」という情報素子の線形配列パターンであり、法律などは印字されたアルファベット（音素を表す文字の体系）という情報素子の線形配列パターンになる。

このあたりから、筆者には少々ついていけない大風呂敷と感じられてしまうが、我慢して次にいくと、物質層と生物・人間層を峻別するある意味で近代的な論理が展開する。

吉田の情報論は「物質」二元論から「物質と情報」三元論へのメタ・パラダイム転換を主張する。「物質と情報」二元論は、厳密に言えば「本源一元論的派生二元論」である。「本源一元論」とは、「質料としての物質」とその「差異／パターン」としての非記号情報」との『不可分の統一体』が物質層の唯一の構成要素であり、生物層・人間層もその物質層を不可避・不可欠の基盤として構築されるといふ自然哲学のことをいう。つまり、「物質とその差異／パターン」二元論である。それに対して「派生二元論」とは、生物層と人間層にのみ固有の「物質と記号情報」との二元性を意味している。「質料Ⅱ形相結合体」という一元論的な物質的自然は、生命の誕生とともに、「非記号情報」から「記号情報」への進化を通じて二元論化する。すなわち、「記号」として機能する質料Ⅱ形相結合体」および「意味」として機能する質料Ⅱ形相結合体」の分化、つまり意味するものと意味されるものとの分化が、発生するのである。

一元論的な物質層から派生する生物層と人間層に固有の二元論の内実は、抽象化すれば大要つぎのようなものである。すなわち「記号」としても「物質の差異／パターン」が認知・評価・指令という三大情報機能ないし三モードの情報機能により、「指示対象」としての「物質の差異／パターン」を指定・表示・制御し、ひるがえって「指示対象」としての「物質の差異／パターン」が何らかの選択過程に媒介されて「記号」としての「物質の差異／パ



タン」を直接・間接に決定するという循環的關係がそれである。具体的事例でいえば、人間の認知・評価・指令的なシンボル性情報が人間的現実を構築し、人間的現実が事前・事後の主体選択に媒介されてシンボル性情報空間を決定するという循環的關係や、生物の遺伝子型が表現型を構築し、表現型が自然選択に媒介されて遺伝子型を決定するという循環的關係がそれである。

このように吉田の言う「情報」というのは人類史的に広大な概念なのだが、とりあえずこういう概念規定と分類を飛ばして、一気に社会科学の方法に触れている部分だけを見る。それは『情報と自己組織性の理論』（一九九〇年）の第五章「社会科学における情報論的視座」のなかの情報論的アプローチの基本的視座について説明している部分である。吉田は、社会科学を貫貫する基幹的アプローチを三つあげ、第一が相互関連的アプローチ、第二が要件論的アプローチ、第三を情報論的アプローチと呼ぶ。以下、吉田の説明はこうなっている。<sup>22)</sup>

第一の相互関連的アプローチは、近代経済学の均衡理論に代表されるシステム論的アプローチをイメージしている。これはまず、相互関連する諸要因（変数）を析出し、その定性的・定量的内容（変数の値）を確定する。次に、諸要因間の相互関連（関数、すなわち変数間の定性的・定量的な対応法則）を定式化する。システムに対する要因間のインプットとアウトプット変換の分析になる。そして、諸要因の均衡の分析になる。この均衡分析がこの相互関連アプローチの究極の課題であり、システムの均衡条件、存在定理、安定条件を吟味することになる。ここでは数学的に、諸変数の関数関係の集合、つまり連立方程式で表現することができる、とする。

第二の要件論的アプローチは、生物化学と心理学と社会科学に特有の、自己保存系（生物・生物社会・人間・人間社会・自動制御機械）を対象にするアプローチである。これも三段階になっている。まず、自己保存の要件

ないし機能的必要、所与の環境下で存続・繁栄・発展するための諸条件を明らかにする要件分析、すなわち価値変数（選好変数、厚生変数、目的変数）の確定作業を行う。次に、自己保存系の内外の各要因が、要件充足をどの程度、促進するか、阻止するかを認定する機能分析である。その帰結は評価関数（選好関数、厚生関数、目的関数）を定式化する作業である。そして要件論的アプローチの究極の課題である許容性分析、つまり要件充足の許容状態の条件、存否、その安定性を吟味する。数学的には、評価関数の集合と内生・外生諸変数の相互連関を表す制約条件の集合とによって示すことができる。例として、効用関数（評価関数）と収支方程式（制約条件）から構成される消費者行動の理論、利潤関数（評価関数）と生産関数（制約条件）から構成される企業行動の理論のミクロ経済理論をあげる。このアプローチを自覚的・系統的に採用したのは文化人類学や社会学の機能学派だが、吉田は「残念ながら、いまだ相互連関的アプローチほどの膨塚をみるに至っていない」が、それと標榜されることなくあらゆる分野で見られるという。

例えば、ダーウィン進化論は、遺伝情報により決定される形態と機能が、たまたま一定の環境のもとで突然変異と自然淘汰が起り、許容範囲の要件充足をもたらす生物だけが存続・繁栄するという、要件論的アプローチの分析であり、心理学の欲求概念は、パーソナリティ・システムの基底的な要件を表したものと考えることができる<sup>(23)</sup>。

最後の情報論的アプローチであるが、これは二つの知見によって支えられているという。

1つは、自己保存系の内外の諸要因とその相互連関が、情報ならびに情報処理によってコントロールされているという知見。生物個体の物質代謝は遺伝情報をもとに酵素の触媒作用を中核とする「高分子性情報処理」によっ

てコントロールされている。人間社会の物質代謝（物質的生産）は、文化情報（制度その他）をもとに経済的意識決定を中核とする「言語性情報処理」によってコントロールされている。個体の刺激—反応過程が「神経性情報処理」によって媒介されていることも知られている。

もう一つの知見というのは、自己保存系の情報ならびに情報処理は、それによってコントロールされる内外の諸要因とその相互連関が、当該システムの要件を一定の許容範囲で充足しうるようなものへと、自然選択ならびに主体選択の選択淘汰作用を通じて変容してゆくという認識である。

情報論的アプローチは、要件の許容充足をもたらす情報・情報処理が結局選択され、それをもたらさない情報・情報処理は結局淘汰される、と把握するのだと吉田は説明している。「情報諸科学の発達は、この「自然選択・自然淘汰」とまさに並置ないし対置すべき「主体選択・主体淘汰」——システム自体による情報の選択淘汰——の思想を、今ようやくくにして確立し始めている」として、情報論的アプローチの根底は「自己保存系の内外の諸要因とその相互連関は、自然選択ならびに主体選択のふるいにかけられた情報ならびに情報処理によって、当該システムの要件を一定の許容範囲で充足しうるような方向へとコントロールされる、という自然認識」にあると述べている。

さらに、情報論的アプローチは社会科学の場合では、次の三つの問題群を抱える。第一が情報処理の「主体」の問題。第二は「シンボル性の情報ならびに情報処理」の問題、第三が情報ならびに情報処理の選択淘汰を規定する条件、すなわち「妥当性」の問題、あるいは「実在」による「情報」の規定という問題である。

ここで少し気になる「妥当性」分析における「実在」とは何か？

それは「情報から独立に存在し、認知・評価・指令の情報機能を通じてコントロールされる対象」のことであり、情報が主体淘汰および自然淘汰のテストないし検証を通じて「実在」に規定されるという側面を表す。この説明はマルクス・エンゲルスの理論に適用できると吉田は説明している。

マルクス・エンゲルスもまた、人間の行為が、表象や観念、意図や目的定立など、シンボル性の情報と情報処理によってコントロールされる事実を鋭く指摘している。ただ周知のように、彼らは、情報現象の〈背後〉にあつてそれを規定する要因——社会的存在——に着目し、〈社会的存在による意識の規定〉を力説した。情報論的アプローチのタームに翻訳するなら、情報ならびに情報処理の〈選択淘汰を規定する要因〉を問題にしたわけであり、情報現象の〈妥当性〉分析を重視したのである。<sup>24</sup>

吉田の三つの基幹的アプローチは、あくまで分析的な立場から分類されたものであつて、研究者が実際の研究で用いるのはこれらを併用ないし統合したものとなつている。吉田にとって、三つのなかでいまだ未発達ながらこれからの科学研究がもつとも力を入れて取り組むべきは、第三の情報論的アプローチになることは当然である。

日本の社会学者の中には、数理社会学を提唱した安田三郎はじめ、社会学の理論を数学的に表現し洗練できると信じる人が何人かいた。吉田民人もその一人である。そのお手本は経済学の均衡理論などで、関数と数式で諸要因の関係を明確に規定し、そこに数値を投入してさまざまな説明を抽出する。これは理論研究であつて、データ集めと集計分析に偏した社会調査でも、現状分析を数量的に扱う計量社会学でもないのだと考える。吉田の壮

大な情報理論もそうした指向が強く出ているが、それにしては『情報と自己組織性の理論』はじめ吉田の残した本の多くには、数式やデータ分析などはほとんど見当たらない。あるいは吉田の数学的研究を、怠慢な筆者が見ていないにすぎないのかもしれないが、あくまで概念づくり、分類操作、理論家に徹した人のように思える。

別のところで、「自己組織性」について、またもや類型を設定した吉田は、最終段階に位置づけているのが「言語情報—内部選択」型の自己組織性であり、これが人間の社会システムになる。自己組織機能はDNA情報による生体の自己組織化から始まり、言語情報による人間社会の自己組織化へという進化をたどる。この進化は、「システムの自由度の増大」という特徴があるとする。

こうした吉田民情報理論を瞥見してみると、「情報」というキーワードですべての現象を包括的に説明できる（説明し直す）のだという情熱と野心を感じるのだが、社会学のこれまでの理論学説史的な観点から見ると、機能主義—システム論の延長上に、進化生物学や遺伝学などの新しい成果を取り込んで、「自己組織性」と「情報理論」を自前で立ち上げた贅力に感じ入る。けれども、この「情報」概念は広すぎて、一般に理解されている「情報」概念に比べ実用的とは思えないし、チョムスキー言語論のように、ことばに繁茂する枝葉を切り落として深層の骨格にまで裸にしてみようのではなく、逆にどんどん枝を広げて伸ばして空を覆ってしまおうとしているかのように見える。

もちろんその中核にシステム論という幹があるのだが、そこにはインターネットの普及がもたらしたAI技術やビッグデータのような問題は、単なる応用問題としてしか現われようがない。選択淘汰で進化するのは、人間の知であって、道具としてのコンピュータやインターネットではなく、仮にAIが自己組織システムになろうと

しても、遠い先のことと吉田は考えていたように、システム論に不勉強な筆者には思える。

情報社会学の開拓者のうち、その後に残された公文俊平の構想では、さすがに情報社会のWEB的進化が視野に入っていて、このような情報社会の進展によって社会に新たな階層分化が生じるという。つまり、この情報を操って望ましい選択や決定に参与する知的エリート「智民」と、それに関与できず情報社会が提供する現実の波間に漂ってそれをただ受容するだけの「痴民」の隔差はどんどん開いていく。最新テクノロジーは人間の生命の根源である遺伝子や生殖を操作可能なものにしつつあるから、もはや空想的段階にはない。

時代が進化するのか、人間の思考が進化するのか、しかしさすがに公文の論は先走っている印象はぬぐえない。さて、気を取り直して当初の社会調査の方法論というローカルな地点に立ち戻ってみると、この情報論については、二〇世紀の実証主義を保守する社会学の社会調査という方法にとって再検討すべき意味を持っているのだろうか？

#### 四 社会調査は生き延びるのか？

一九世紀の後半に、フランスなどで国家や地方政府が「統計」というものを定期的にかくさん取るようになって、一種のブームが起った。それまで単なる漠然とした噂や印象、大雑把な推測でしかなかった「社会の全体像」が、数字によって示されたと人々は思い、統計によってさまざまな社会現象が正確に捉えられると信じたが、当時の統計は現在のものに比べれば調査項目も限られていて、捕捉率も悪かった。それでも、国民国家が完成して

官僚組織が整備されるに伴って、統計は進化していった。それは自然科学で定着した現象の記録に数字を用いるという態度が科学の制度化によって、一般にも普及していったことを表していた。

二〇世紀に入ると多くの国で、国勢調査をはじめ経済、財政、人口、衛生、民生など行政が基礎とするデータを統計に求めるのが当然のようになった。デュルケーム学派の社会学が、統計を活用して当時のフランス社会の諸側面を「社会的事実」として研究したことはよく知られている。統計とは基本的に数量データであるから、数学的操作が可能である。本稿の最初でも触れた推測統計学も、巨大化する大衆社会を正確に捉えようとすれば、確率論的な技法を応用していく必要があったわけである。

しかし、統計の作成は大規模になればなるほど、手間もかかりお金もかかるから、国家や自治体が必要に応じて限られた項目を調査する他はない。そこに入らない現象は統計から漏れているから、別途、小規模でも独自に調査をする必要がある。社会学などは、さまざまな社会現象を対象として研究を行うから、多くの問題で利用できる統計は限られる。独自のデータが必要であれば、自前で調査を行わなければならない。そこで、社会学の教育には社会調査の技術が必須となるようになっていった。

そしてコンピュータが発明され、第二次大戦後、それが大量のデータ処理を可能にし、ハード面での能力を飛躍的に高めるのと並行して、プログラムの開発や、小型軽量化が進み、パソコンが廉価になって普及するなかで、数量的な統計処理の技術は文系の研究者にも容易く扱えるようになったことは、いまさら言うまでもない。しかし、その頃はまだ、それはあくまで数量データを計算処理する道具であって、ほとんどの統計データの収集は調査機関の人間、あるいは特定の研究者が、しかるべき場所で時間を費やして手配り目配りして拾っていく努力が

必要だった。

社会学にとっても、必要なデータを手にするには、みずから社会調査を設計し対象者を想定し、調査票を作り、フィールドに出て苦勞して集めた回答を、コンピュータにかけるためにカテゴリー・データを数量化する手間をかけていたのである。経済統計のような金額で生産量や投資額や貿易額や給与水準が測定できる場合、あるいはデータが人口や事業所数や従業員数のような絶対量を示す数値ならば、そのまま的な尺度で計算ができるが、人の行動や意識を測定したデータは、せいぜい順位尺度までしか捕捉できない。その際も、標準的なやり方では、調査票という道具に質問文と選択肢を設け、選択肢に番号を振っておけば、データはいちおう数字でカウントされる。それをコンピュータで集計解析する技法もさまざま開発された。

しかし、一方で社会学が扱う問題には、そのような数量化では捉えきれない言語的表現がどうしても登場する。あなたはこの問題にどのような意見をお持ちですか？ といった質問に回答を求める場合、選択肢を四つ用意して、単に「1 賛成、2 やや賛成、3 やや反対、4 反対」というような形で答えてもらうのであれば、数量化してコンピュータで集計できる。必要なら他の変数との相関も出すことができる。しかし、なぜその意見に賛成するのかをもう少し詳しく説明してもらおうと、四択ではわからない「質的」要素が、言葉で語られることになり、それをアフターコーディングで無理に数量化することに抵抗を感じる社会学者が現われて、このような方法への異議と不信を唱えるだろう。

そこで、一九七〇年代以降、社会学のなかでいわゆる言語派と総称される社会学者は、数量統計的な社会調査を捨てて、ことばをそのままデータとみなして分析しようとする別の道を歩き始めた。それをここでは論じる余



裕がないが、彼らは社会と人間を捉える手段として、統計数字ではなく人が語ることばをそのまま採集し、分析できると考える。それは実は、今に始まったことではなく、自然科学をモデルとする実証主義、論理主義に対抗する、社会学のなかの別の流れとして数字ではなく言語を重視する立場に繋がっている。

しかし、ここで触れたいのは、それとは別に、二一世紀になって思わぬところからデータ収集と言語という問題が現われたことである。

筆者が本稿であえて、チョムスキーの言語論が補助線になるかもしれないと書いたのは、この問題がいま新たに従来の方法論への再考を促すかもしれないと考えたからである。それは、コンピュータがもたらす技術と環境が、いまや情報処理の道具の域を超えて、インターネットの進展によってデータ収集のあり方自体が、変わってしまう事態が現われたと思われるからである。

ビッグデータとAIの話題で一般に引き合いに出される例は、コンビニなど小売店での人々の消費行動のデータである。カードを使って商品を購入すれば自動的に、その商店、時刻、買った商品、金額が記録されている。現金で買った場合も本人認証はメであるが、購買行動の記録は残る。あらゆる商店で膨大な記録が日々積み上がっている。それは当初は、小売店側の売上げや仕入れや経営のための情報として使うためであっただろうが、スマホや電子マネーの普及で、そのデータは各店舗やチェーン本部に保存されるだけでなく、クラウド・ネットワークで共有されることも可能だ。コンビニには防犯用のカメラも設置されているだろうから、訪れた客の画像も残る。こうして日々あちこちで記録されたビッグデータを誰がどのように使うのか、巷の消費者にはよくわからない。

社会調査では、調査協力者には調査の趣旨をよく説明してラポールを確保し、その個人情報には漏れないように秘匿しなければいけないと教えているが、いまやそんな調査など不要なほどに、われわれの生活は知らないうちに記録され蓄積されどこかでデータとして情報処理されているわけである。もし、特定個人に関する情報が検索されるなら、プロフィールの網にかかり、自分がいつどこでどういう行動をしていたかが、知らないうちに見知らぬ誰かに把握されている事態は、いまおそらく実現しつつある。

ということとは、調査をする側から見ると、今までのような素朴で人間的な社会調査はますます困難になる一方で、しかるべきビッグデータのありかさえ確保・利用できれば、面倒な社会調査などやらなくても、大量のひとの行動は全件処理できる、標本調査のような調査より、はるかに大量で細かなデータが使えるようになる、あとはデータ処理の速度と効率だと考えるエンジニアはたぶんとくさんいるはずだ。

しかし、筆者はやはり、ちよつとまてよ、そんなにことは単純ではないと考え直す。ビッグデータと一括して呼ばれる多様で大量のさまざまなデータの何をどう使うのか、それは無闇に面白半分で任意にコンピュータにかけて相関をとればいいわけではない。使うべきデータを選んで組み合わせるのは、やはりもの考える人間なのであり、コンピュータではない。

そのために、もう少しビッグデータの応用の事例を見る必要がある。そこで、ここではひとつのきわめて実用的なある研究を参考にとりあげたい。

工学系の災害研究、林良嗣・鈴木康弘らのグループが行った「レジリエンス（しなやかな回復力）」をテーマとした共同研究のうち、「ジオ・ビッグデータによる地震災害リスク評価とレジリエントな国土デザイン」（執筆

者：柴崎亮介・秋山祐樹・加藤孝明）が採用している手法をみてみよう。<sup>(26)</sup>

この研究の意図は、以下のような発想から出発している。

二〇一一年三月一日に発生した東日本大震災を契機に、日本各地の自治体では、地震被害想定や危険度に関する調査を行い、その結果を公開・提供し、地域住民や民間企業などが閲覧・利用できる環境作りをすすめている。しかし、その情報の多くは自治体ごとに作成され公開されていて、町丁目単位やメッシュ単位で集計されたものしか公開されない。調査の方法や被害予測の基準も自治体間で必ずしも同じでないため、地域間の比較が難しい。そこで柴崎らは、「建物一棟一棟が見える細かさで、しかもそれが自由なスケールで集計可能で、都道府県・市区町村の壁がない状態で日本全国、スケールシームレスに被災状況が分析・推定ができる空間情報プラットフォーム」を実現することを考えた。

具体的には、国勢調査などの様々な公開統計情報、緯度経度座標付き電話帳データベース、デジタル住宅地図などの一般的に利用可能で、しかも日本全土をカバーできる様々な統計・空間データを用いて、地震災害のリスク（より具体的には地震による建物破壊と火災のリスク）と、災害への初期対応力を評価するための建物一棟単位のマイクロな空間データ（Ⅱマイクロジオデータ）の基盤整備を行ない、それらを用いて地震災害リスクと初期対応力の簡易的な評価手法を提案し、日本全国を対象に適用することで、地域間の相対的な地震災害リスクと災害対応力の可視化を実施したという。

ここで注意したいのは、ビッグデータというけれども、その中身は実に多様で、それぞれは別の目的、別の手法で収集されたものであることだ。全国の建物情報は、GPSの航空からの位置情報もあれば、人が歩いて一軒

一軒表札を確認した地図情報や映像記録もある。そこに住む人間の属性まではさすがに特定していない。そこは人口統計や建築記録などから推定することになるが、それは常時変動している。多種多様であるがゆえに、そのどれを選び、どのようなアルゴリズムで組み合わせる分析し、その結果をどう読むかは、それらの意味を判断できる専門家という人間の能力にかかっているのである。コンピュータは自己組織性を創造するような学習する人格ではなく、将棋名人の棋譜をなぞってはループして修正している機械にすぎない。

大地震という予知不可能な自然現象に対して、理系の工学的発想は、空間的・地理的情報をできるだけ詳細にして、そこに統計的数値を組み合わせ、どこまで意味のある数字を高性能のコンピュータによって出せるかだけが、中心的な関心である。

柴崎たちの手法は、デジタル住宅地図から日本全国の建物約六千万棟の位置情報を取得し、ポイントデータ化する（建物ポイントデータ）。これで建物の位置、面積、階数、用途が観察できる。そのデータに対して統計情報やミクロな空間データの情報を配分して、一棟一棟の災害リスクと初期対応を計算できる環境を実現したという。

建物一棟一棟には図1に示す情報が付加される。

一棟ごとのリスク情報は三種ある。それは次のようなものになる。

●被災リスク情報 ・地震後に火災が発生し消失する可能性（火災リスク）

・地震の揺れによって建物が倒壊する可能性（倒壊リスク）

●被災に対する初期対応情報 ・周辺の消防組織による出火建物の消火力（公助力）

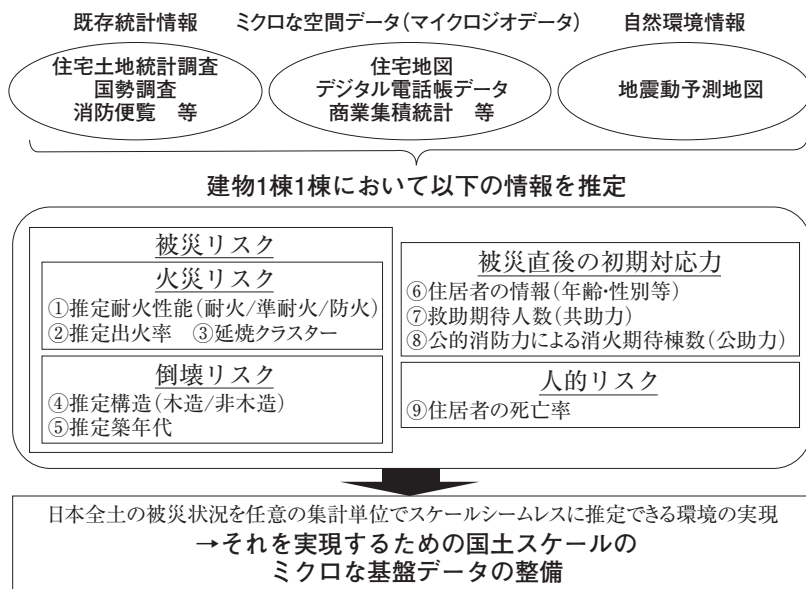
・地域住民によって期待される倒壊建物からの救助力（共助力）

●人的リスク情報 ・そこに居住している住民が被災する可能性（人的リスク）

既存統計情報、マイクロジオデータ、自然環境情報といったデータ・ソースから、火災リスク、倒壊リスク、被災直後の初期対応力、人的リスクの四つを推定し、日本全土の被災状況を任意の集計単位でスケールシームレスに推定できる環境を実現するため、ミクロな基盤データを整備するというものである。

建物ポイントデータに対して、さまざまな属性情報を連続的に付加していくことで、最終的には建物一棟一棟に地震による被害に関する多様な情報が付加される。それらを任意

「統計的」社会調査法とビッグデータ



出典：林良嗣・鈴木康弘編著『レジリエンスと地域創生』明石書店、2015、p.177

図1 建物単位のマイクロジオデータの整備の流れ

の空間単位（例えば町丁目や学区、メッシュなど）で集計することで、任意の地域で被災リスクと、災害への初期対応力の計算が可能になる。また人的被害から初期対応力を差し引くことで、その地域の最終的な人的被害の推定が可能になる。集計単位が高精細なため、被災リスク、初期対応力、人的被害想定の結果を任意の地域間で定量的に比較評価することが可能になる。

この研究の結果、ジオ・ビッグデータにもとづくさまざまな被災シミュレーションとハザードマップが作成され推定値が計算されている。

さて、これはビッグデータといっても、防災、大地震への対処・復元（復興）というひとつの目的に合わせて選ばれたものである。データ・ソースはいろいろあるけれども、それを建物一棟単位で詳細に組み合わせることが、すでにあるデータを使えば可能になったからこそできる作業である。ここにAI技術が使われたのかどうかはわからない。いずれにしても、ビッグデータの適切な使い方は、このようなものなのだろう。日本列島のどこにでも大地震・大災害の起こる可能性があり、それを正確に予測できない以上、既存のデータでできるかぎり詳細に分析しておくことは、有効な作業である。それが各地の住民に活用されれば、防災対策として成果をあげることができる。

しかし、これは本稿の前半でみたビッグデータ論議とは、どうも違う次元のものに筆者には思える。

AIとビッグデータの現状をバラ色の夢として語る人々を「サイバー・リバタリアン」と呼び、シンギュラリティ（技術的特異点）を目標に世界をサイバー空間として呑み込む動きは、実は新自由主義の「勝ち組のユート

「ピア」なのだと言う本山美彦のような立場もある。<sup>(28)</sup>

社会学の一般向け啓蒙書としてベストセラーを出すことで知られるランドル・コリンズは、コンピュータと人工知能の進化についてこんなことを言っている。

もし人間の知能は社会的であるということが正しければ、これらの創造的思考のそれぞれを基礎づける特定の種類の社会的相互作用があるということになる。たとえば、新たな科学的理論をつくり出すのは、他の科学者たちと相互作用をしている特定の科学者たちである。事実、一部の科学者たちが創造的な「ホットスポット」に位置し、それ以外の人たちはありふれたルーティンワークに従事する周辺の位置にあるようなネットワークがあり、それについては多くのことが知られている。同様のことは、作曲家のネットワークや文学者のネットワークについても当てはまる。社会的世界には多くの異なった領域がある。知能の社会学理論の重要な要素となるのは、それゆえ、ある個人型の人たちとの関係においてどこに位置づけられるかということが彼または彼女の思考を決定することになる、という考え方である。ここで、若し作曲の出来る、あるいは小説の書けるコンピュータをつくりたいのであれば、作曲家あるいは小説家のネットワークの中にそのコンピュータを置き、それらの人たちと相互作用する能力をそのコンピュータに与える必要がある。そうすれば、コンピュータは、作曲するにせよ小説を書くにせよ、課題を遂行するための能力と動機づけの双方を獲得することになるだろう。

もちろん、すべてのことを一挙に行うようにわれらのコンピュータをプログラムすることはできない。最

初は単純なことからはじめ、だんだんより複雑なことへと築き上げていかねばならない。基本としてまず構築しなければならぬのは、人と相互作用の出来るコンピュータである。次にそのコンピュータが単独でも考えられるようにしなければならぬ。つまり、外部で他の人びとと交わす会話を内部にもち込み、「心のなかで」沈黙の会話ができるようにしなければならない。私たちは、このコンピュータが誰と話したいのかを自分で認識できるようにしなければならないし、また人びとが時間を割いてこのコンピュータに話しかけることに興味を抱くような方法を考えなければならない。要するに、このコンピュータを、社会的世界を生き抜いていく普通の人間とまったく同じようなものにする必要があるのだ。もしそうならば、ある種の観念や概念をその他の観念や概念よりも重要なものとする方法を私たちは手に入れることになり、それゆえこのコンピュータは会話をしていない状態の時に考えたいことについてプレファランス選好をもつようになるだろう。そしてそうならば、社会学が示唆するように、私たちはすべてのことができるコンピュータをもつことになるだろう。つまりそれは、日常のゴシップやジョークから高度に創造的な思考まで、人間的な能力の範囲を構成するすべてのことができるコンピュータとなるだろう。

そのようなコンピュータ・プログラムはまだ存在しない。<sup>20</sup>

この皮肉交じりの文章は、一九九二年に書かれていたから、その段階では人工知能にはできないことが多かった。電子計算機が情報貯蔵庫になり、通信機器になり携帯電話と合体した二〇世紀はとくに終わり、いまはコンピュータが述べるような、ある分野の有能な人間と交わってディープラーニングすれば、作曲をしたり小説を書く



ようなコンピュータができてくるのは当然のようである。おまけにインターネットとAI技術は、二〇世紀には考えられなかったような膨大な人間の行動をほとんどまるとカウントし記録するところまで来た。

しかし、その先にあるものをわれわれはバラ色に語り過ぎていないか？

## おわりに

こうしてコンピュータ・サイエンスにも情報論にも素人である筆者は、現代の人工知能研究や「ビッグデータ」の先端が向かう先についても、理解できた範囲でそれが人間の知的活動として推進されることに今のところ、格別の異議はない。近代以降の必然的な知の流れからすれば、それは当然追求されて然るべき目標だと思う。しかし、それを人類の限らない幸福を実現するテクノロジーの楽園のように語るには欺瞞だと思うし、それを明るくポジティブに語る人の無邪気な主張は、結局偏狭な歴史認識、文化的偏向に傾いてしまうと思う。二〇世紀の社会科学が探求してきたことは、およそ六十億人に増殖した地球上の人間が、日々作りなしている社会の現実について、とりあえず対処すべき操作可能な問題について、どこまで正確な現状認識とそれを説明する理論、そこから未来にむけて効果的な対策を立てられるか、ということに尽きる。そのために手段としての社会調査もあるのだと思う。

だとすれば、最初のささやかな問い、現代のテクノロジーが推進するAIとビッグデータ利用という知的活動に、社会調査の方法は技術的のみならず思想的に折り合いをつけられるのか、という論点をあげて本稿の結びと

したい。

チョムスキーは、言語を獲得した人間が脳の活動機能によって達成できる知の限界を、生産的な普遍文法の「創造性」に求めた。しかし、それがコンピュータ・サイエンスそしてウェブ社会にそのまま利用されたわけではなく、むしろそのような演繹的な普遍主義、論理主義の徹底ではなく、「AIとビッグデータ」論議は、ディープラーニングのような、チョムスキーが嫌った経験的・帰納的方法に回帰していると言えるのかもしれない。

ともかく社会調査は今までの方法を後生大事にやっていたらいいのか？技法上の改善や統計ソフトの改良程度のテクニカルな話で終るなら、それはそれで安心であるが、筆者にはいまのところはつきりした見通しが見えない。「ビッグデータ」のAI分析の活用例を見る限り、考えられることが二点ある。

一つは、従来の数量統計的な社会調査の方法で調査を実施するのは、現実的に困難を抱えているということ。信頼性を確保した精密なデータを得ようとすればするほど、相手が人間であることが立ちはだかる。回収率は落ちる一方である。かといって膨大なビッグデータが使えればそれでよい、などとは到底言えない。

もう一つは、数字でコンピュータにやってもらえることはどんどんやってもらえばいいし、われわれも使いこなす能力はつけなければならぬが、社会学が扱うデータには数量化しにくいもののがかなりあることだ。ビッグデータの多くは数字になっていないもので、たとえば画像はデジタル情報になっているが、ツイッターの文などはなんらかの操作で変換しなければならぬ。キーワード検索や機械翻訳などの技術がさらに進展したとしても、ことを数字に置き換えればよい、とは思えない。チョムスキー言語理論は、そこに大かきな手がかりを与えてくれることを期待されたけれども、どうもうまくいかなかった。

われわれが世界に起きた出来事と、人間の行動や意識を正確に把握しようとするとき、普遍主義・論理主義をとるか、経験主義・実証主義をとるかは、研究の態度として大きく違うのだが、どちらかが勝利を収めるというようなものではない。要するに、使ってみてうまくいけば使うのである。

## 註

- (1) 竹内啓「ビッグデータと統計学」〔現代思想〕二〇一四年六月号）青土社、二九〜三〇頁。
- (2) 本山美彦『人工知能と21世紀の資本主義 サイバー空間と新自由主義』明石書店、第二章。
- (3) 『EJ Times Japan』日立製作所の矢野和男氏に聞く：/http://eeimes.jp/ej/articles/1607/14/news027\_2.html
- (4) 文部科学省『平成28年版科学技術白書』四二頁。
- (5) 西垣通『ビッグデータと人工知能』中公新書、二〇一六年、二〇頁。この部分の原典はビクター・マイヤー＝ショーンベルガー＋ケネス・クキエ（斉藤栄一郎訳）『ビッグデータの将来』邦訳書、九八頁。
- (6) 西垣通『ビッグデータと人工知能 可能性と畏を見極める』中公新書、二〇一六年、三六頁。
- (7) 西垣通『ビッグデータと人工知能』中公新書、二〇一六年、三二〜三七頁。
- (8) 西垣通『ビッグデータと人工知能』中公文庫、二〇一六年、三七〜四〇頁。
- (9) 西垣『同書』四三頁。『演繹』と『帰納』のほかに『仮説推量』がある。『仮説推量』とは、ある一般ルールと個別事実とから、個別条件をみちびくというもの。ある一般ルールの存在を前提にして、この事実が起きた背景にこの条件が成り立つのではないか、という仮説を作る。『人間は死ぬ』と『ソクラテスは死ぬ』とから、『ソクラテスは人間だ』と推量するわけだ。だが、仮説推量は必ずしも成り立たない。もしかしたら死んだのはソクラテスという名前のイヌかもしれない。ビッグデータ分析にもついでに行動をおこすとき、（例外のある）帰納だけでなく、（間違いもある）仮説推量の処理が行われている、という点はよくよく注意しないといけない。」

「統計的」社会調査法とビッグデータ

「統計的」社会調査法とビッグデータ

- (10) Quételet, Lambert Adolphe Jacques "La physique sociale" 1869, ケトレー『人間に就いて 上』岩波書店、一九四八年。  
彼は人の社会的データのみならず身体的データについても研究を行っている。特に人の身長に対する理想的体重と実際の体重を比較する指数、つまりボディマス指数(ケトレー指数)を提案し、これは公衆医学上も重要な貢献となった。
- (11) 竹内啓「ビッグデータと統計学」『現代思想』二〇一四年六月号、二九〇―三二一頁。
- (12) 大黒岳彦「ビッグデータの社会哲学的位相」(『現代思想』二〇一四年六月号)、一三四―三五頁。(大黒岳彦『情報社会の(哲学)』勁章書房、二〇一六年、八三―八五頁に再録)
- (13) N・チョムスキー『統辞構造論』福井直樹・辻子美保子訳、岩波文庫、二四〇―二五五頁。
- (14) N・チョムスキー『言語と精神』(川本茂雄訳) 河出書房新社、三七頁。
- (15) Chomsky, Noam. *Cartesian Linguistics*. New York: Harper and Row, 1965. N・チョムスキー『デカルト派言語学——合理主義思想の歴史の一章』(川本茂雄訳)「新版」みすず書房、一九七六年、四三―五七頁。
- (16) 田中克彦『チョムスキー』岩波現代文庫、二〇〇〇年(原著一九八三年) 五七頁。
- (17) F・ソシュール『一般言語学第一回講義』相原奈津江・秋津伶訳、エディット・バルク、二〇〇〇年。
- (18) 西垣通「解説」(田中克彦『チョムスキー』岩波現代文庫、二〇〇〇年) 二六五―二六八頁。
- (19) 三浦雅士「あとがき——インターネットとポストモダン」(三浦雅士編『ポストモダンを超えて』平凡社、二〇一六年。四三五―四三八頁。
- (20) N・チョムスキー『統辞構造論』福井直樹・辻子美保子訳、岩波文庫、九頁。
- (21) 公文俊平「情報社会学への招待」(公文俊平編『情報社会学概論』N T T出版、二〇一一年) 一―二頁。
- (22) 吉田民人「情報と自己組織性の理論」東京大学出版会、一九九〇年。
- (23) 吉田民人『同書』一九九〇年。一三一―一三三頁。
- (24) 吉田民人『同書』一九九〇年。一三五頁。
- (25) 公文俊平「情報社会のいま——新しい住民たちへ」N T T出版、二〇一一年。
- (26) 林良嗣・鈴木康弘編著『レジリエンスと地域創生 伝統知とビッグデータから探る国土デザイン』明石書店、二〇一五年。ジオ・ビッグデータを扱っているのは、第五章一節と二節、一七六―二〇二頁。

- (27) 柴崎亮介・秋山祐樹「マイクログジョデータベースによる地震災害リスク評価」(林良嗣・鈴木康弘編著『レジリエンスと地域創生』明石書店、所収論文) 二〇一五年。一七八頁。
- (28) 本山美彦「人工知能と21世紀の資本主義 サイバー空間と新自由主義」明石書店、二〇一五年。
- (29) ランドル・コリンズ『脱常識の社会学』第二版、井上俊・磯部卓三訳、岩波現代文庫、二五〇～二五七頁。Collins, Randall "Sociological Insight: An Introduction to Non-Obvious Sociology" 1992 by Oxford University Press inc.

#### 参考文献

- 大黒岳彦「情報社会の〈哲学〉 グーグル・ビッグデータ・人工知能」勁草書房、二〇一六年。
- 岡嶋裕史「ビッグデータの罠」新潮社、二〇一四年。
- 公文俊平編『情報社会学概論』N T T出版、二〇一一年。
- 公文俊平「情報社会のいま——あたらしい住民たちへ」N T T出版、二〇一一年。
- 盛山和夫『統計学入門』ちくま学芸文庫、二〇一五年。
- フェルディナン・ド・ソシユール『一般言語学第一回講義』相原奈津江・秋津伶訳、エディット・パルク、二〇〇八年。
- 竹内啓「ビッグデータと統計学」(『現代思想』二〇一四年六月、第四二卷九号 青土社)
- 田中克彦「チョムスキー」岩波現代文庫、二〇〇〇年。
- 西垣通「ビッグデータと人工知能 可能性と罠を見極める」中公新書、二〇一六年。
- 西垣通、ドミニク・チェン「情報<sup>ヒューマン</sup>は人を自由にするか」(『現代思想』二〇一四年六月、第四二卷九号 青土社)。
- 西垣通『ペシミスティック・サイボーグ』青土社、一九九四年。
- Chomsky, Noam. *Cartesian Linguistics*. New York: Harper and Row, 1965. N・チョムスキー『デカルト派言語学——合理主義思想の歴史の一章』(川本茂雄訳)「新版」みすず書房、一九七六年。
- ノーム・チョムスキー「チョムスキー言語基礎論集」福井直樹編訳、岩波書店、二〇一二年。

「統計的」社会調査法とビッグデータ

「統計的」社会調査法とビッグデータ

- ノーム・チョムスキー『生成文法の企て』原著一九八二年、福井直樹・辻子美保子訳、岩波現代文庫二〇一一年
- ノーム・チョムスキー『統辞構造論』福井直樹・辻子美保子訳、岩波文庫、二〇一四年。
- ノーム・チョムスキー『言語と精神』（川本茂雄訳）河出書房新社、二〇一一年。
- 林良嗣・鈴木康弘編著『レジリエンスと地域創生 伝統知とビッグデータから探る国土デザイン』明石書店、二〇一五年。
- 三浦雅士編『ポストモダンを超えて』平凡社、二〇一六年。
- 本山美彦『人工知能と21世紀の資本主義 サイバー空間と新自由主義』明石書店、二〇一五年。
- 文部科学省『平成二八年版 科学技術白書』二〇一六年。
- 吉田民人『情報と自己組織性の理論』東京大学出版会、一九九〇年。
- 吉田民人『社会情報学とその展開』勁草書房、二〇一三年。
- ランドル・コリンズ『脱常識の社会学』第二版、井上俊・磯部卓三訳、岩波現代文庫、二五〇～二五七頁。Collins, randoll "Sociological Insight An Introduction to Non-Obvious Sociology" 1992, by Oxford University Press inc.