

インタビュー調査のデータ分析に関する一考察

廣瀬 幸市*

*心理講座

An essay on the data analysis of the interview survey

Koichi HIROSE*

**Psychology Course in Department of Training Programs for Educational Support Professionals, Aichi University of Education*

要 約

人文社会科学の研究領域のうち、対人援助を旨とする心理臨床学の分野においても近年は質的研究が盛んであり、最近ではテキストマイニングを用いた研究も散見されるようになってきている。混合研究法やコンピュータを用いた内容分析などの手法は、隣接分野の看護研究に比べて普及はまだであるが、シンギュラリティ問題が喧伝される中、その潮流の行方は予測できない。本稿では、コンピュータによる自然言語処理の加速度的処理能力の増大によって益々便利になったテキスト型データ分析を用いた研究の様相をサーヴェイしながら、臨床心理学分野において盛んな質的研究あるいは隣接分野における混合研究の在り様を比較しながら考察した。

臨床心理学分野の質的研究では、グラウンデッド・セオリー・アプローチを特に取り上げて、そのコーディング作業を検討した。その結果、コーディング作業の進展に伴ってデータから理論を生成するプロセスにおいて、データと概念の距離関係を最適化していく中で意味生成が非連続的に得られること（解釈の第2ダイナミズム）が判明した。隣接分野の混合研究では、「グラウンデッドなテキストマイニング・アプローチ」の手法を特に取り上げて、人間がテキストを理解する方法とコンピュータがテキストを解析する方法がどれほど異なっているのか、について考察を深めた。また、テキスト型データ分析を行う上で、母語とする言語の違いによって思考様式（世界認識の在り方）が異なるという点を考慮に入れる必要があることを指摘した。

Keywords : テキスト型データ分析 グラウンデッド・セオリー・アプローチ 混合研究法 コーディング

0. はじめに

人工知能への憧憬はコンピュータが生まれて間もなくの時代から近未来SFのテーマとして描かれ続けていたが、近年の囲碁や将棋界におけるコンピュータ・ソフトの快進撃を見せられるようになって、一般人にもシンギュラリティ（技術特異点）が意識されるようになってきて、これまで無意識にあった“人工知能に人類が乗り越えられる恐怖”がようやく水面に見え隠れするようになったと思われる。

人文社会科学の領域においては、人間が繰り広げる思考空間や対人社会的世界の多彩で魅力的な性質・可能性に包まれていて、シンギュラリティがもたらす問題は2045年問題として産業界で喧しいものの、学問領域としては一部の領域で2045年問題を踏まえた議論がなされる以外にはその余波を被ることは少ないように思われる。近年、人文社会科学の調査研究でもテキストマイニングの手法およびその研究利用が広がっ

てきて、コンピュータによる自然言語処理の加速度的処理能力の増大によって、やがて質的研究においても一種の“シンギュラリティ”が起こるだろうとの潜在的な願望あるいは期待に基づく論調を嗅ぎ取れなくもない。まだ意識に上らないような潜在的課題を引き上げて論じるのは時期尚早と思われるかも知れないが、本稿ではその辺りを問題意識の底として、現在興隆しているテキスト型データ分析を用いた質的研究のゆくえを考察してみたいと考えている。

1. コンピュータを用いたテキスト型データ分析

近年ではコンピュータを利用したテキストマイニングが進んでいる。テキストマイニングは、予めコンピュータに設定されている自然言語処理機能、統計解析機能やコーパス参照機能などを用いてテキスト型データを自動的に演算処理する手法で、インターネットの発達に伴う膨大な電子テキストの活用によって近年飛

躍的に処理能力が増大している(黒橋, 2015)。大隅(2006)によると、人工知能研究の主流としてのデータ・マイニングが登場し、それと言語学研究、自然言語処理研究が融合して、テキストマイニングが生まれたという。これら諸研究の融合体、各種の技術要素、種々の方法論の集合体と考えられるのだという。ここでテキストとは、電子コード化された文字情報のこと、即ちテキスト型データを指している。そして、関連する研究分野は次のように多種多様である。主要なものだけ列举すると、「自然言語処理あるいは計算機言語学」、人工知能、「情報検索、情報処理」、計量言語学、「言語学、社会学、行動科学など」、「記号論、テキスト論、カテゴリー論、意味論など」である。これらの関連研究分野で用いられる方法論については、主なものだけでも次のようなものがある。即ち、パターン認識、各種統計的手法(特に多変量解析、多次元データ解析)、分類手法(特に判別分析、クラスター分析)、文書管理情報処理技術などである。

ところで、樋口耕一はテキスト型データの分析方法「計量テキスト分析」を提案して、その方法を実現するためのフリーソフトウェア「KH Coder」を開発・公開してきた。「計量テキスト分析」とは、計量的分析手法を用いてテキスト型データを整理または分析し、内容分析(content analysis)を行う方法である(樋口, 2006)。この分析の実践においては、自然言語処理や統計解析などの近年のコンピュータ技術を活用する手法となっている。2001年以降「KH Coder」により文章型データ活用を普及・促進させ、学術分野に止まらず実務分野の利用者を広げてきた。KH Coderを利用した応用研究は現時点で、学会発表と論文・書籍を合わせると1,500件に達している(樋口, 2017)。今後はただ件数を増やすのではなく、KH Coderが一層うまく利用されて優れた応用研究が生み出されることを企図する段階にある、と彼自身も認識しており、どのように利用すればデータから社会的意義のある発見につながり易いのかを探索している。そこで、KH Coderを用いたテキストマイニング手法である「計量テキスト分析」を取り上げて、実践・経験知にどのように迫れるのか、見ておこう。

計量テキスト分析がスタートした時点では、社会学の社会調査法の分野で質問紙調査における自由回答項目を分析するのに用いられた。当初は新聞記事、雑誌記事、宗教教典等が分析対象となっていたが、その後にインタビュー・データが加えられて(川端, 2001)、現在ではテキスト型データ一般を扱うようになっていく。川端の研究が進められていた当時は「コンピュータ・コーディング」の名称で呼ばれ、コンピュータを用いてコーディングを行うことに重点が置かれていたが、次第に量的方法による探索の側面が重視されるようになり、計量テキスト分析という名称が考案された

(川端, 2003)。KH Coderでは共起ネットワーク、対応分析を用いてコード間の結びつきを探り、コンコダンス機能、文書検索機能などを用いて計量的分析の意味するところを元のテキストに戻って確認できる。計量テキスト分析では、テキスト型データの言葉を文脈から切り離して機械的に数え上げることに利点を見出しているが、それらの利点の1つはデータ探索の観点からであり、もう1つは分析の信頼性向上の為である(樋口, 2017)。しかし一方で、研究者が文書中のテキストを直接読んで質的に解釈することを否定はしていない。むしろ、計量的な分析によってデータの中で注目すべき部分を探した上で、その部分を研究者が質的解釈を行う方法も認めている。これは川端(2003)の提案した計量テキスト分析の考え方に基づいていて、樋口(2006)はこれに依拠しつつ内容分析の考え方を拡充したものであると考えている。つまり、質的方法と量的方法とを循環的に用いることを推奨している(樋口, 2014)。

また、内容分析の考え方を自然言語処理・統計分析・コンピュータを活用することで、分析実践に活かし易くなるのではないかと考えた彼は、内容分析を行うための具体的手順・方法の一種として計量テキスト分析を定義したのだが、これによって、計量テキスト分析は単なるデータの記述(description)ではなく推論(inference)を含むという、内容分析の考え方を受け継ぐことを明示することになった。内容分析の分析対象はコミュニケーション内容であり、単に内容を記述するだけでなく送り手の意図や置かれた状況を推論し得る。内容分析の扱うデータはシンボリックなものであり、あくまで間接的に対象や概念を表すと考えられ、1つのメッセージを送り手にとっての意味や、受け手にとっての意味というように、複数の意味を推論することが可能である。したがって、計量テキスト分析を適用したデータはどのように研究に役立つか、と考えていくことは、各種データを基にどのような社会的状況を推論できるのか、ということになる。

ところで、質的データを量的に扱える形に変換する作業において、データを幾つかのカテゴリーに分類していくコーディング作業が行われる。このコーディングの分類基準(コーディング・ルール)を作成する営みは非量的作業であり、量的分析の結果がこの過程に依存している。コーディング・ルールを作成する作業は、研究者の社会的想像力が発揮されるべき極めて重要かつ質的な作業である(樋口, 2006)。研究者の持つ理論的仮説や研究目的が異なれば、同じデータに向き合っても異なるコーディング・ルールが作成されることになる。このような科学的方法としての厳密性の維持に関して、KH Coderでは、コーディング・ルールを開示して信頼性を維持するという考え方に依拠している。大量のテキスト型データも計量テキスト分

析において KH Coder を利用して計算すれば、結果として様々な図表を得ることができるが、その中でどの部分が重要なのかを判断して、その部分を基に推論を行うのは研究者の役割である。計量テキスト分析ではコーディング・ルールを開示することによって、研究者がデータ内容を基に推論を行った過程を辿れるように、信頼性 (dependability) を確保している訳である。このようにして、研究の成否を左右する重要なプロセスである、データ内容に基づく推論はコンピュータによって自動化できないプロセスでもある (樋口, 2017)。

以上に見てきたように、テキストマイニングの主流を占めるとされる計量テキスト分析は、コーディング・ルールの作成においても、また、KH Coder を用いた計算によって得た図表などの活用においても、研究者がどの部分が重要なのかを判断し、それを基に推論を行うプロセスが大切なのであり、その過程はコンピュータによって自動化できないということが分かった。

また、本稿では紙幅の関係で取り上げることができなかったが、田中 (2017) もテキストマイニングに関して、自然言語処理は汎用的な基礎解析だけでなく課題に即した技術開発等が行われ、その際に言語を大胆に捨象して近似することが日常的に行われるが、その結果として現在の技術水準では文脈などの大局的な言語情報は失われて、得られる知識断片には専門家による関連知識の補完や解釈が必然的に求められる、と指摘している。

2. GTAに見られるテキスト型データ分析

人文社会科学の領域における研究は、総じて人を対象とする研究一般となるので、研究の方法論に関して言えば、量的なデータ解析に持ち込める数量的データに変換するには限界がある。これまでの既存の先行研究により、量的データ解析でめざましい結果が示せるような洋々たる分野は既に研究され尽くしている感がある。そのような背景もあってか、一度は廃れた様相を呈していた質的研究が近年になって再び脚光を浴びて、新たな装いで様々な方法論が咲き乱れるようになった。

このような状況の下、とりわけ臨床心理学の分野で広く用いられるようになった質的研究が幾つかある。グラウンデッド・セオリー・アプローチ (以下、GTA と略す) である。本論では、後に論じる計量テキスト分析と対比するために、GTA に見られる質的データの分析に関わる考え方を押さえておくことにしたい。

人類学のフィールド・ワーカーであった佐藤 (2008) は、質的データ分析を①重層的な文脈の解明、および②現場の言葉と理論の言葉のあいだの往復、という意味で、殆どあらゆるタイプの質的研究には「一種の翻訳作業としての性格がある」と論じている。彼

は、この質的研究の性格を米国の人類学者クリフォード・ギアツが提唱した、「分厚い記述」と「経験近接的概念」という2つのアイディアを援用して、質的研究の本質的特徴について理解を進める説明概念を提供している。本稿では、論の進め易さから、後者のみ見しておくことにしよう。

Geertz (1973/1987) の言う「経験近接的概念 (experience-near concept)」とは、調査対象となる人々 (就労者、患者、地域住民等) が日常生活で見たり、感じたり、考えたりした内容をその人々自身の言葉で表現する際に使用する用語や概念を指す。一方で、「経験遠隔的概念 (experience-distant concept)」とは、調査者や分析者が何らかの理論的枠組みを踏まえて使用する概念やそれに対応する用語を指す。これを分かり易く言い換えれば、前者を「現場の言葉」、後者を「理論の言葉」と名づけることができる、と言う (佐藤, 2008)。

質的研究として認められるためには、研究対象となる人々の意味世界を構成する現場の言葉を、研究者コミュニティの意味世界における理論の言葉に移し替えていく作業が非常に重要であり、これによって他の事例や出来事とのより厳密な比較が可能となるだけでなく、個別具体的な出来事や事例の理解を越えて一般的なパターンや法則性を割り出していくことが出来るようになる。後に見ていくことにするが、木下の修正版 GTA では、分析焦点者を設定することにより、社会相互作用を読み込もうとしている。

現場の意味世界のリアリティを生かしながら、それを抽象的な理論の言葉に置き換えていく作業を行うためには、一定期間のトレーニングと実際のデータ分析の体験の積み重ねを踏まえた地道な努力が不可欠となる (佐藤, 2008)。その際、この作業の重要なカギとなるのが、GTA においては「コーディング (qualitative coding)」と呼ばれる手続きである。テキスト型データの分類を行うコーディングの種類として、オープン・コーディング (open coding)、軸足コーディング (axial coding)、選択的コーディング (selective coding) がある。

また、テキスト型データの処理および収集に関して、継続的比較分析 (constant comparative method)、理論的比較 (theoretical comparison)、および理論的サンプリング (theoretical sampling) がある。これらの分析手段を通して、現象から収集された具体的記述 (テキスト型データ) を抽象度の高い概念に段階的にまとめていく。この一連の分析の結果については、現象の構造とプロセスを図式化した「概念図 (conceptual model)」および、現象を簡潔に説明する「ストーリー・ライン (storyline)」のセットで提示されることになる。このような最終形 (現象の説明様式) に到達した一連の分析プロセスが、提示され

た結果の反証可能性を保証していると考えられている。我が国で広く用いられるようになっているストラウス・コービン版GTA（以下、SC-GTAと略す）は「データ収集→テキスト化→切片化→オープン・コーディング→軸足コーディング→選択的コーディング→ストーリー・ライン作成」のステップで構成されている（木下, 2008）。これに対し、SC-GTAと二分する利用者を獲得している修正版GTA（以下、M-GTAと略す）は、分析手順の開示性と明示性が保たれる範囲内で可能な限りの手続き簡略が行われている。SC-GTAと異なり、切片化がなく、コーディングでもオープン・コーディングと選択的コーディングしかない。テキスト型データに根差しながらも、テキストの或る部分の意味だけを検討するのではなく、他の類似例をも説明できるかどうか考えるという、2つの検討を行っている分析者の意味の選択的判断を積極的に分析的プロセスに取り入れることによって、深い解釈に至ろうとするからこそ、このようなコーディングの方法論を採用している（木下, 2003）。

SC-GTAとM-GTAのいずれでも、GTAにおいてコーディングという作業は、オープン・コーディング、軸足コーディング、選択的コーディングと作業を進めるに従い、段階的に概念生成から概念間関係の検討へと移行していく（住, 2014）。ところが、M-GTAでは理論的メモ作成において、概念生成の局面から概念間関係の検討が開始され、概念・定義・カテゴリーの各々同士の関係性により全体を説明することが目指されている（木下, 2003）。つまり、GTA一般がデータから1次コード（ラベル）化し、そこから要約して2次コード化、さらに包括的な3次コード化と進むように、コーディング作業においてデータから間接的に離れていく階層的なまとめ方をするのとは異なる（木下, 2007）のである。M-GTAでは、grounded-on-dataで分析していき、あくまでデータの着目部分から言える範囲で解釈していくのだが、概念を生成する際には一定程度の多様な具体例が説明できるかどうかをチェックしながら、データと概念との距離バランスを取る。そして、類似例が増えるごとに定義に照らし再検討し、必要に応じて概念の修正を行う。このような作業を繰り返して、一定数の多様な具体例と概念との関係を最適化していくのであり、それはデータと概念との間を行きつ戻りつしながらの作業プロセスである。この作業中、データの解釈、その解釈からデータへの確認、それを交互にしながらデータに最もフィットする概念に仕上げていくことが重要とされていて、1つ一つの作業が解釈と連動して進められている（木下, 2007）ことが特長である。

それゆえ、M-GTAでは解釈を行う分析者の役割が重要であり、そこで要とされているのが「研究する人間」の見識である。「研究する人間」には、主観と客

観との往復を通して徹底的に自問自答を繰り返すことが求められ、最終的に社会に発表する結果に対する責任を持つという研究姿勢にも掛かっている。このような重層的な文脈の読み込みこそが特定の研究領域の研究者にしかできない作業であることは、自ずと想像されるところである。

また、解釈を行うことに内包される重要な知見に関して、木下（2007）が指摘する2種類のダイナミズムが大変参考になる。解釈の第1ダイナミズムは、オープン・コーディングの段階でデータの具体例と概念との間で起こる対応関係のことに関わっている。これに対して、第2ダイナミズムは具体例から類似例へ、次に概念と概念との比較へ、そしてカテゴリー間比較へと、抽象度を上げていく中での比較検討を行っていく作業の途中に起きるものに関わっている。今、注目したいのは第2ダイナミズムの方である。

木下（2007）は、最初の提唱者であったグレイザーがストラウスを批判して言った分析手続きを例に引きながら、意味生成について論じている。すなわち、ストラウスのように事前に考えておいた概念的枠組みをデータに押し付けて解釈するのではなく、データを忠実に見ていけばその意味は「浮上（emergent）」してくるという、グレイザーが再三強調した考え方についてである。ここで、浮上するというのは、分析がまとまってくる、あるいは柱となる解釈がはっきりしてくるということである。木下（2007）は、これは分析作業をしていて非連続的に着想が得られることだと考えた。つまり、発想法やabductionとも関係することとして、先に見たような比較のレベルを上げていく中で、まとめになるような着想が非連続的に浮かぶのではないかと説明している。その上で、第2ダイナミズムの特徴を、非連続なものであること、基礎的作業より抽象度を或る程度上げたところでの比較から得やすいこと、実際の作業手順に隠れて作動するので他者には分かりにくいこと、経験的に確認するしかないこと、と纏めている。さらに、この時に得られたアイデアが研究者にとってリアリティ感のある独自の内容であることも、特徴として付け加えている。

上のような指摘から、量的であれ質的であれ、研究において最も重要な点に分かるように思われる。即ち、GTAによるテキスト型データの分析においては、データの具体例から概念を生成し、概念間関係からカテゴリーを生成し、全体の概念図やストーリー・ラインを作成していくプロセスの途上で、分析者の行う解釈の中に非連続的な着想の浮上が生じることがある、という重要点である。木下（2007）は、ここが質的データ分析の醍醐味であり、コーディング作業においてはオープン化から収束化への「質的移行」の分岐点でもある、と言っている。データに密着した分析から独自の理論を生成する質的研究法であるGTAでは、分析の真

際作業に隠れて解釈の第2ダイナミズムが作動していることは、分析を行う研究者でなければ経験できない。そのため、厳密な作業手順を推し進めていけば“自動的に”着想が“浮上”するのではないかと期待しがちである。しかし、これまで本稿で確かめてきたように、現場の言葉と理論の言葉の相互翻訳(佐藤, 2008)、等、の思考プロセスにおける懸隔が目に見えない形で存在していて、如何に演算処理能力を増大させても辞書的あるいは定型的に広く一般的な文例を収集したコーパスを利用してもコンピュータ自動翻訳の作業では乗り越えることは困難であろうと想像する。

と言うのも、佐藤(2015)も述べているように、翻訳ソフトによって自動化されるのは、基本的には語単位ないし文単位での置き換えというルーチン的な部分だからである。これに対して、専門の翻訳家がたとえば外国語の作品の翻訳をおこなう場合には、単語や文章という単位だけでなく、少なくとも次にあげるようなさまざまなレベルの文脈を考慮に入れていかなければならない。すなわち、特定の文章が埋め込まれている前後の文章の脈絡、段落を単位にした文脈、段落と段落の関係、特定の段落が含まれる節あるいは章の関係、書物全体レベルでの文脈(全体のプロット、テーマ、文体等)である。このような重層的な文脈の解明および現場の言葉と理論の言葉の往復、あるいはそれらのプロセスを踏まえた脱文脈化・再文脈化の作業は、質的研究の全プロセスにおいてどうしても機械任せにすることが出来ない「アート」として要素を多分に含んでいる。このアートとしての側面はテキスト型データの読み込みの作業においては必須の要件である。

3. 混合研究法におけるテキスト型データ分析

質的研究にコンピュータ利用を積極的に推奨する研究者である稲葉・抱井(2011)は、構成主義的GTA(Charmaz, 2006/2008)にKH Coderによる計量テキスト分析を組み合わせた混合研究法として、「グラウンデッドなテキストマイニング・アプローチ(以下、GTMAと略す)」を提案している。彼らは、研究者自身によるテキストの読み込みや分析によってテキストの可能性を展開するdivergentな方向性と、コンピュータのテキストマイニングによって知見を収束させるconvergentな方向性の2つの作業を繰り返すことにより、テキストデータをマイクロとマクロの両方から把握することができるだけでなく、テキストデータの表面には現れていない深いレベルでの仮説生成や、新たな知識構築ができる可能性がある、としている。彼らはGTMAの研究評価規準を検討する中で、研究者自身が行う質的データ分析とGTMAとの比較検討を行っている。そして、confirmability(確証性)で最も大きな差が出るとした上で、credibility(信憑性)とtransferability(転用可能性)はGTMAを用いたから

と言って簡単に改善される訳ではないと述べている。そうではあっても、GTMAを用いることで研究者自身が行う質的データ分析だけでは見えてこない矛盾・疑問点を見つけ出すことができ、credibilityやtransferabilityのより充足した分析結果を導き出せる可能性があるとしている。

なお、dependability(確実性)については、GTMAではコンピュータによる言語解析・多変量解析などを用いた機械的処理による分析のため、分析結果に研究者による「解釈の飛躍」が入り込む余地が少ないとして、従来型質的データ分析より担保されやすいと称揚している(稲葉・抱井, 2011)。dependabilityに関しては、計量テキスト分析でも見たように、研究者の持つ理論的仮説や研究目的が異なれば、同じデータに向き合っても異なるコーディング・ルールが作成されるのだが、それをどのように考えるかは研究者の立場によって異なる。例えば構成主義的GTAにおいては、それは本質であると捉えられる(末田・抱井・沖潮, 2016)。dependability以外のことは混合研究法(Creswell&Clark, 2007/2010)のメリットと同質であり、大量のテキスト型データを処理しなければならない研究の場合においては異論の余地はない。

しかしながら、そのテキスト型データのリソース元がインタビュー調査によるものであった場合、インタビューが半構造化面接の方法を取ったとしても、インタビュアーの与える影響を無視できない対話となる以上、インタビュイーの回答部分をトランスクリプトしたデータであっても、対話の文脈を読み込んだデータ解釈となっているかどうかにおいて、差が出てくるのではないだろうか。つまり例えば、研究者自身が行う質的データ分析において生成されるコードと、GTMAに基づく統合的分析で採用されるコードとは異なっている場合が少なくなく、それもコーディングの分類の発想自体から異なることになるのではないだろうか。稲葉・抱井(2011)も「人間がテキストを理解する方法と、コンピュータがテキストを解析する方法は根源的に異なっている」と述べており、「研究者がテキストを読み込むことで生成される印象、コード化・カテゴリー化の結果として見える重要なポイント、コンピュータによる可視化という3つの結果の間には、何らかの矛盾が生じる可能性が高く、そこから研究者にとっての様々な疑問点が浮かび上がってくる可能性も高い」ことを認めている。しかし、それに対しては、「このような矛盾点や疑問点についてはメモ書きを通して記録・考察し、それらを解消するために、改めて原文を参照し、コードやカテゴリーを再検討し、テキストデータの可視化を再度試みるという作業を繰り返すことで、テキストで語られている現象に近づくことができる可能性がある」と述べるに止まっている。本稿では、彼らが止まる領域を越えて考察を進めてみ

たい。

ところで、このことで思い出されるのは、計量テキスト分析を初めに提唱した川端 (2003) の次の視点である。「分析の対象とする文章を読んで、文字通りに解釈するのではなく、文をいったん語に分解し、その語と語の間の連関の強さを元に単語の使われる暗黙の意味構造を新たな『潜在的論理』として、いわば文法のように取り出し、データ理解の枠組みとして用いるのである。つまり計量的な分析が、多変量解析の手法を用いて潜在的な概念を見いだすのと同じようなことが、質的データを対象として実現可能となるはずである」。つまり、ここで彼が多変量解析において見出されている「潜在的な概念」に準えて表現している「潜在的論理」のことである。彼自身、コンピュータ・コーディングの欠点として、非あいまい化の問題、文脈理解の問題を挙げている。しかし、KH Coder に連なる計量テキスト分析の初期には、かなり大量の質的データに自動的にコードを振る作業を優先する方略を採用して、文脈や曖昧化の問題を犠牲にしている。そこを犠牲にしても、樋口 (2004) の言う Correlational アプローチのメリットを活かして、分析者の持つ理論や問題意識の影響を極力受けない方向性を指向している、と言えるだろう。このようにしてコンピュータによる一連の多変量解析を通して把握されたテキスト型データ全体の「潜在的論理」とは一体何を意味するのだろうか。また、それは現在、人文社会科学の領域でも用いられている、因子分析における因子の命名といったような類の活用で済むものなのだろうか。本稿で課題にしたいのは、「人間がテキストを理解する方法と、コンピュータがテキストを解析する方法は根源的に異なっている」(稲葉・抱井, 2011) という、その在り方である。

現在、コンピュータによる自然言語処理の飛躍的能力向上で、語義曖昧性解消・構文的曖昧性解消・「構文・各解析」・「照応・省略解析」・談話構造解析の精度が上がってきているが、それは一文単位のことであって、文章間となると十分な精度ではないようである。我々の日常的な言語使用法をコンピュータ言語に置き換えることによって行う自然言語処理が掴まえてくる、単語間の連関の強さを基に見出された「暗黙の意味構造」とされる「潜在的論理」とは、人間がテキストを理解することで到達できる意味構造に読み替えが利くものなのだろうか。それとも、コンピュータがテキストを解析することで掴まえた意味構造は、そこへ至るプロセスが根源的に異なるが故に、人間の言語使用による意味構造の理解に対して何らかの一考あるいは刷新を求めてくる類のものなのだろうか。

筆者にこのような問いを迫るのは、本稿冒頭の2045年問題もさることながら、棋界における藤井六段の目覚ましい活躍が念頭にある。彼がコンピュータ

将棋プログラムソフトを相手に指し手研究を行っていることは広く知られている。プログラムソフトとの棋戦ではなく、いったん人の頭脳を通した彼独自の創造的な指し手との棋戦であるからこそ、あれだけ実力のある上位の棋士を相手に互角以上の戦いが繰り広げられるのだらうと思われる。テキスト解析に擬えるのはカテゴリ・ミステイクかも知れないが、コンピュータ・プログラムソフトの解析する指し手を前に、人間の棋士は近年敗退を余儀なくされる場面が増え、人間の思考方法そのものの再考あるいは拡張を求められているように見えることと同様の事態が密かに進行しているのではないかと疑わざるを得ない。時代的な背景がそうさせていたに過ぎないかも知れないが、稲葉・抱井 (2011) も川端 (2003) もこの問題に深入りすることなく、GTMA あるいは計量テキスト分析を用いて混合研究を行うことで研究者自身による質的データ分析は補完される、と論じるに止まっていた。

この疑問に現時点で完全に解答することは不可能だと思われる。その代り、研究者は各々の信じたい認識論的立場を選択することが許されているのだらう。今後の研究に期待したいところだが、本稿では次のような立場で一応の回答にしておこうと考えている。それは次のようなものである。

M-GTA におけるコーディング作業は、QDA ソフトウェアを用いても、結局は研究者 (分析者) の“主観と客観の往還”が作業のプロセスの途上で行われており、文法に基づき定型文などを参考に修正を施したコーディング手順を自動化した自然言語処理で辿るプロセスとは、決定的なところで“似て非なるもの”とならざるを得ない。ここに働くメカニズムが先述した「解釈の第2ダイナミズム」である。これは文法上の手続きを超えたところで、研究者の脳裏に閃くダイナミズムだからである。それゆえ、人工知能が破竹の勢いで驚異的な進化を遂げている中、音声対話システムで会話が支障ない程度に、文脈解析や談話構造解析すら乗り越えて、意味構造の抽出までもが出来そうに思われる節があるが、筆者はそうなるとは思えない。

4. 臨床心理学研究におけるテキスト型データ分析

前節のように考えるのは、筆者が心理臨床学研究という特殊な領域に身を置いて質的研究を行っている事情と深く関係しているように思われる。心理臨床学研究は元々、生身の相談を主として対話を通して解決しようとする来談者との関わりを研究素材にしてきた。かつて事例研究においては高度に圧縮された記述をテキストとしていたが、近年の質的研究においては主としてインタビュー調査により収集された言語データをテキストとして利用することが多くなっている。そのテキスト型データを、例えばGTAのような研究方法によって一連の質的分析を施して、同じ様に来談者と共に

問題解決に取り組む研究者仲間に研究成果を提供し、その心理臨床実践活動に資する研究としている。このような研究内容に関して、とりわけ心理臨床学研究の性質として特筆すべきことは、セラピストクライアントの二者間相互作用による影響の不可避性および、二者間で交換される会話の文脈依存性の高さを挙げておきたい(廣瀬, 2015)。この対話で交錯する遣り取りは、文法に載せた文の意味内容だけに止まらず、文章の行間とも比喩されるような意味空間によって伝達される内容が含まれている。このような意味内容の分析においては、文章を文字通りに受け取るのでは十分でない。二人の関係性に載せたところで伝達が可能となるような意味内容が読み取れることも必要である(西平, 2001)。そこは佐久川(2013)が「対人支援の現象学は必然的に解釈学という性格を持つ」と言うように、既に解釈の世界に入っているところだが、そのような遣り取りを目的とする領域の人文社会科学であることを改めて指摘しておくことは無駄ではないだろう。

臨床心理学の隣接領域である医療援助者、とりわけ看護の研究分野では早くから質的研究が展開しており、質的アプローチだけでなく量的アプローチも合わせた混合研究法が広がっている(Creswell&Clark, 2007/2010)。コンピュータを用いた内容分析やテキストマイニングを活用した研究も増えていて、近年では看護における暗黙知を解明する研究が徐々に増えてきているという(上野・川野, 2015)。心理臨床実践を研究しようとするならば、とりわけその実践知や経験知を伝えるための研究であるならば、暗黙知を取り扱わない訳にはいかない。筆者もかつて考察に及んだことがある(本多・廣瀬, 2016)が、本稿では紙幅の関係で触れることができない。心理臨床学研究において「暗黙知」は今後益々重要な意義を持つことになるが、それについては別稿に改めたい。

4. テキスト型データにおける言語による影響

コンピュータ言語と人文社会科学で用いられる自然言語との相違を別の角度から照射してみよう。情報システムの領域でユニークな研究を続けている金田重郎は、これまで要求分析手法に関する論文を数多く発表してきているが、その中に日本語の特性に注目した研究が見られる。中でも特に、河合隼雄の「中空均衡構造論」に着目した論考(金田, 2010)は大変興味深く、本稿ではそれを参照しておきたい。

彼によると、日本民族は縄文時代を通じて文字を持たなかったところへ、中国(漢)から漢字が輸入されたが、抽象概念の発達が不十分であったことから、その後漢字を本来の字義とは異なって用いて日本語の抽象語を創造するという工夫を行った。それで結果的に、日本古来の概念を表現しようとするならば、漢字

の概念のように単独で表現することはできず、組み合わせることによって新しい抽象概念を創造して表現しようとする。表現したい「和の心」に一つの単語を当てるのではなく、その周辺を中空均衡構造のように日本語単語(複数)で貼り付けて全体として「意味」を分かってもらおうとするのだ、と言う。これを彼は「日本語表現の多視点性」と呼び、後の論考において、認識論の観点から英語と日本語の言語の違いに着目して情報システムにおけるクラス図や概念データモデリングの理解促進を提案している(金田・井田・酒井, 2014)。

彼は、印欧語の英語と日本語の認識の相違を、金谷(2002)の主張を引用して「神の眼と虫の眼」と指摘した上で、「三上章による日本語文法」に基づいて日本語に適したGTA(J-GTA)を提案している(金田・永田, 2011)。GTAのラベルやカテゴリーに名詞を用いて作業することが日本人には必ずしも使い易いものではないと考え、プロパティとディメンジョンの見直しを提案している。彼はその手法の効用を測定するために或るシステム評価に適応して従来型(ストラウス・コービン版)GTAと比較してみたところ、処理時間の向上効果は限定されていたという。彼はその原因を、プロパティ・ディメンジョンの作成、理論的飽和を目指すための比較の作業に時間を要するためだと分析している。その後の彼の研究には残念ながら、J-GTAの改良型提案等は見つけられなかったが、筆者の視点からすると、考えられることが2つほどある。一つは、金田の用いている従来型GTAがストラウス・コービン版であることである。グレイザーは後にストラウスと大論争の果てに袂を分かったが、その争点は手続きの厳密さにあったと言われている(木下, 2007)。ストラウス・コービン版はグレイザー版ほどではないが、プロパティやディメンジョンの作成手続きがしっかりしており、手間が掛かる。さらに理論的飽和のための比較の作業を厳密に行えば、当然ながら作業の短縮は期待したほどには見られないだろう。木下のM-GTAを使用したなら、結果は多少変わったかも知れないと思われる。もう一つは、主題の抽出の問題に関わることである。金田(2011)が着目したように、主語優勢言語である英語より主題優勢言語である日本語では構文情報から容易に抽出できる可能性が高いが、GTAの作業手順に則って行う場合、KJ法のように直感的にラベルを選ぶ方法と比較して、その作業時間は期待するほど短縮できないだろう。彼は情報システムにおけるクラス図や概念データモデリングの研究を重ねてきているので、コンピュータ言語による手続きに則って工夫を積み重ねてきており、人間の思考過程において生じる直感等の思考ループを極力排除しているだろうと想像される。

本稿で注目しているのは、日本語による抽象概念な

どの表現法として中空均衡構造のような思考様式をしていること、また英語などの印欧語族の言語と比較して、世界の認識様式がそもそも全く異なっていて、そのような認識に基づいた言語相互交流を行うことを通して自身の心の在り様を表現して伝えようとしている、ということである。改めてこのようなことを確認したことによって、人文社会科学領域の研究テーマをインタビュー調査によって言語的にデータ収集しようとするならば、当然のことながら、日本語特有の思考様式に基づいた言語使用になっていること、それ故、その話し言葉をトランスクリプトして得たテキスト型データはインタビュー過程の文脈の中で意味を考慮しなければ理解しがたいことになっているだろう、ということが容易に想像されてくる。勿論、トランスクリプトしたテキストは、一文毎に見れば文法的に理解可能ではあるが、それらが複数掛かってようやく表現しようとされていたインタビューの伝えたかった「和の心」(金田, 2011)については、現時点の AI 技術では理解できるレベルにまで追いついていないだろう。

以上、紙幅の関係でここまで議論を止めるが、日本語を母語とする日本人の世界認識様式まで見据えた日本語理解とその活用によるテキスト型データ分析への目配せにより、テキストマイニング活用においては、従来通りの日本語文法に基づいた自然言語処理による知識発見を超えたところで、根本の思考プロセスの違いや発想という「解釈のダイナミズム」を考慮に入れる必要があることが確認できた。大隅(2006)の指摘するように、非構造的な自由記述文のテキストマイニングを行うには全く異なる視点からのアプローチが必要なこと、「現状のテキスト型データの数量化・軽量化を通じて知識発見を行う方法だけではなく、“何か別の道”があるだろう」こと、発話者・発言者(回答者)の“言いたいこと、述べたいこと”を拾い上げるような「仕組み作り」を考えること、が大切になって来ているのである。

5. おわりに

本稿では、2045年問題に象徴されるシンギュラリティが人文社会科学の領域においても起こるのだろうか、という問題意識を背景にして考察を進めてきた訳だが、その問いへの回答は研究者自身の認識論的立場により異なるだろう。自身の研究分野でどのような研究法が用いられるようになるのかという潮流に加えて、どのようなリサーチ・クエスチョンを立てるかという研究姿勢がより一層問われてくるものと思われる。

人文社会科学の研究、とりわけ心理臨床領域における質的研究の場合、コーディング作業において SC-GTA のように直輸入的に名詞をカテゴリーやラベルに用いる方法ではなく、M-GTA のように文章のまとまりからテーマの主題を抽出していく工程が有効であるこ

とを指摘する声が論文作成現場には多く、実際に大学院修士課程の修了要件である修士論文に関しては紀要関連書に現れる数以上に多いという感触がある。

計量テキスト分析を始めとするテキスト型データ分析は数百以上のサンプルを扱うのに対して、修士論文を始めとする GTA による質的研究が扱うのはせいぜい数十のオーダーであることを勘案しても、サンプルとする質的データの取り扱いに関しては、「整理」の側に寄ったテキスト型データの活用と「分析」の側に寄ったテキスト型データの活用とを意識的に使い分けて、樋口(2014)の言うように「量的分析と質的分析とを循環的に用いることでデータへの理解を深める」ことを心掛けていくことが必要であると考えられる。

謝辞：本研究は科学研究費助成事業(基盤研究(C)15K04120)による研究成果の一部である。

【引用文献】

- Charmaz K (2006) *Constructing Grounded Theory*. Thousand Oaks : Sage. (抱井尚子, 末田清子 監訳 (2008) *グラウンデッド・セオリーの構築*. ナカニシヤ出版.)
- Creswell JW & Clark VP (2007) *Designing and Conducting : Mixed Methods Research*. Sage Publication. (大谷順子 訳 (2010) *人間科学のための混合研究法*. 北大路書房.)
- Geertz, Clifford (1973) *The Interpretation of Cultures*, New York, NY: Basic Books (吉田禎吾・柳川啓一・中牧弘允・板橋作美 訳 (1987) *文化の解釈学 I・II*, 岩波書店.)
- Glaser B & Strauss A (1967) *The Discovery of Grounded Theory*. Chicago : Aldine. (後藤隆, 大出春江, 水野節夫 訳 (1996) *データ対話型理論の発見*. 新曜社.)
- 樋口耕一 (2006) 内容分析から計量テキスト分析へー継承と発展をめざして, 大阪大学大学院人間科学研究科紀要, 32, 1-27.
- 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.
- 樋口耕一 (2017) 計量テキスト分析および KH Coder の利用状況と展望, 社会学評論, 68(3), 334-350.
- 廣瀬幸市 (2015) 事例研究法と質的研究法—事例を資料とするときに留意すること, 臨床心理学 増刊第 6号, 116-122.
- 本多祐子・廣瀬幸市 (2016) 「臨床に求められる知」に関する一考察 — エーリッヒ・フロム思想と知識創造理論を通して —, 愛知教育大学教育臨床総合センター紀要, 6, 27-35.
- 稲葉光行・抱井尚子 (2011) 質的データ分析における

グラウンデッドなテキストマイニング・アプローチ
の提案：がん告知の可否をめぐるフォーカスグル
ープでの議論の分析から, 政策科学 18(3), 255-
276.

- 金谷武洋 (2002) 日本語に主語はいらない, 講談社
- 金田重郎 (2010) 「中空均衡構造論」に基づく情報シ
ステムの要求分析に関する一考察, 研究報告情報シ
ステムと社会環境(IS), 113(4), 1-8.
- 金田重郎・永田健 (2011) 日本語特性に着目した日本
語 GTA(J-GTA)の提案, 電子情報通信学会技術研究
報告. SS, ソフトウェアサイエンス 111(168), 1-6.
- 金田重郎・井田明男・酒井孝真 (2014) 日本語仕様文
からの概念モデリングプロセス—英語7文型と関数
従属性に基づくクラス図の作成—, 研究報告情報シ
ステムと社会環境(IS), 128(3), 1-8.
- 川端亮 (2001) コンピュータ・コーディングによる宗
教的ライフストーリーの記述, 宗教と社会 7, 133-
153.
- 川端亮 (2003) 宗教の計量的分析：真如苑を事例と
して, 大阪大学大学院人間科学研究科 平成14年度
博士論文.
- 黒橋禎夫 (2015) 自然言語処理, 放送大学教育振興会.
- 西平直 (2001) 哲学と臨床心理学. In : 下山晴彦・
丹野義彦 編：臨床心理学とは何か. 東京大学出版
会, 213-228.
- 大隅昇 (2006) テキスト型データのマイニング—最近
の動向とそれが目指すもの—. In : テキスト型デー
タのマイニング. テキストマイニング研究会, 1-28.
- 戈木グレイグヒル滋子 (2008) 質的研究方法ゼミナ
ール：グラウンデッドセオリーアプローチを学ぶ,
医学書院.
- 佐久川肇 (2013) 質的研究のための現象学入門 第2
版. 医学書院.
- 佐藤郁哉 (2008) 質的データ分析法—原理・方法・実
践, 新曜社.
- 佐藤郁哉 (2015) 質的データ分析の基本原則とQDAソ
フトウェアの可能性, 日本労働研究雑誌 57(12),
81-96.
- 末田清子・抱井尚子・沖潮満里子 (2016) 構成主義的
グラウンデッド・セオリー・アプローチを用いた分
析事例：二つの視点・二つの理論, 青山国際政経
論集 (96), 25-57.
- 住政次郎 (2014) GTA 入門：言語データを質的に分析
するには, 竹内理・水本篤 編著 外国語教育研究ハ
ンドブック【改訂版】：研究手法のよりよい理解の
ために, 松柏社.
- 田中省作 (2017) 自然言語処理からみた「テキスト」
と「テキストマイニング」, 社会学評論, 68(3),
351-367.
- 上野栄一・川野雅資 (2015) テキストマイニングの基