

「生成 AI のハルシネーション」の社会学

—実験臨床社会学の視角から—

檜田 美雄

摂南大学現代社会学部

kashida.yoshio@nifty.ne.jp

An Sociology on "Hallucination of Generative AI"

: From the Perspective of Experimental Clinical Sociology

KASHIDA Yoshio

Setsunan University

Keywords: Generative AI, Hallucination, Experiment, Co-evolution

要旨

本論文は“生成 AI のハルシネーション（幻覚）”という現象に注目する。生成 AI という非生物が「ハルシネーション」（という知覚と認識のずれ）を起こすことはほとんどなさそうなことなのだが、にもかかわらず人々は「ハルシネーション」が AI に起きていると考え、いささか強引にその程度の大小で生成 AI を評価し、その原因を考え、そして生成 AI の運用方法の工夫で「ハルシネーション」を減らそうとしている。とするのならば、この苦労と工夫は興味深い社会現象であり、その解明は科学技術の社会学の魅力的な研究対象である、ということができるだろう。本論文は、この「生成 AI の科学技術社会学的解明」を「実験臨床社会学」（檜田、近刊）の立場から行おうとするものである。

我々はまず 1 節で、人工知能（AI）を論じることがどのようにして社会学になり得るのかを素描する。ついで 2 節で、人々が「生成 AI が引き起こすハルシネーション」をどのように語っているかを中島（2024）内の記載を元に検討する。第一に、人々が「正しい結論」を出力する生成 AI を強く求めていること、第二に、その一方で、出力の 100% の正しさを期待することが困難であるので、最後は人間の判断力に頼るしかないとも考えていること、この二種の事実が判明した。

我々はその次の 3 節で、実際に生成 AI がどのような挙動（入力と出力の対関係）をするかを、檜田（2025a）を元に確認した。大学のレポート課題に答えることを求められた生成 AI は、単独でユーザーが「おかしい」と判断できる出力（類型 1, 2, 3 の「ハルシ

ネーション」) もしていたが、それだけでなく、いろいろ組み合わせて考えることで「おかしい」と判断される出力(類型4の「ハルシネーション」)もしていた。

つぎの4節では、まず今回の実験臨床社会学が、大学教育における「生成 AI 闊利用問題」にも有効な視座を提供するものだろうという予想をしたうえで、後半においては、2節と3節を組み合わせての考察を行った。すなわち、第一に人々の「ハルシネーション」をなくそうとする方法は、いずれも「ハルシネーション」の生じ方と矛盾していたり、生成 AI そのものの有効性を削ぐ方向のものだったりするため、有効性に疑義があること。第二に、生成 AI に対して「出力されたハルシネーションには原因があるはずだ」と考えている現在の我々の思考の仕方は、我々が生成 AI の「ハルシネーション」の祖型であると考えている精神科領域の「ハルシネーション(幻覚)」について、当事者が「このハルシネーションには原因があるはずだ」と考えた時の考え方と同型であること、この2点が発見された。

さいごの5節では、議論のまとめと残された課題についての言及を行った。

また、本論文の末尾には「資料：レポート課題におけるハルシネーションの事例」を置いた。

1 「我々の社会の知の状況」を照らし出す文化社会学の素材としての「AI 論」

「人工知能(AI: 以下では AI を用いる)に関する議論」がかまびすしい。しかし「AI に関する議論が大量に交わされること」と、「AI についての議論から矛盾や混乱が減っていくこと」は違うことである。

なぜそんなことが起きるかといえば、「AIに人々が期待していること」は大変に多様で、かつ、その「多様な期待」が重なり合って、「AIの実際の作動メカニズム」とは関係のない議論のフィールドが作り上げられてしまっているからだ。

したがって、ここに社会学の探究の対象が存在することになる。すなわち、「AIについての人々の期待」と「期待に基づいた AI への働きかけ方」を丁寧に記述することで、(「AI の本質や可能性がわかる」わけではないが)「人々がどのようなことを AI に求めているのか」とか、「人々が自分達の期待を果たさない AI に対して、どのようなやり方でどのような改革をしようとしているのか」ということが分かってくるのである。

このように構想できる AI の社会文化的研究のうち、「AI のハルシネーション(幻覚)に関する研究」は特に興味深い。もちろん、生成 AI に対してハルシネーションを起こしていると主張することは、端的にカテゴリー錯誤(Ryle 1949=1987)であるといえるだろう。「幻覚」とは(少なくとも精神医学においては)「対象なき知覚」(中根晃 1975→1985: 160-1)だからだ。生成 AI には感覚器がなく、したがって、不適切な出力を生成 AI がしたときに、その原因を感覚器での知覚にかかわる情報処理系の問題に還元することなどできるはずもない。

けれども、我々は「生成 AI がハルシネーションを起こしている」といまや普通に発言している。少なくとも、そこには「ハルシネーションを起こさない AI であってほしい」という願望があるといえるだろう。おそらくは、生成 AI が「ハルシネーション」を起こしているという判断は、「AI 的知性への期待」の反作用として存在しているのであろう (Caluori 2023)。とするのなら、「『我々の社会の AI 理解』の現況を照らし出す文化社会学の素材として『AI のハルシネーションに関する観察結果』」を分析していくという研究構想が成り立つことになる。

2 AI におけるハルシネーションとは何か/それはどのように語られているか

ところで、AI における「ハルシネーション (幻覚)」とは何か¹。

手元にある生成 AI 解説本 (中島 2024) には、「嘘を出力する」という節があり、そこには次項で引用している内容が書かれている。

2.1 中島 (2024) におけるハルシネーションの解説はどうなっているか (1)

以下、中島 (2024) の「嘘を出力する」の主要部を引用することから、この節をはじめよう。

生成 AI が嘘を出力する現象は「ハルシネーション (Hallucinations)」と呼ばれ、とくに注意したいポイントです。

ハルシネーションが起こる原因のひとつは、学習したデータに間違った情報が含まれている (1) ことです。……インターネットの情報には誤った情報や偏ったデータが存在するため、結果としてハルシネーションにつながります。

また、文章生成 AI は次の単語を予測しているに過ぎないため、学習していない情報を求められた場合、あたかも真実のように嘘を出力する (2) こともありえます。

……

生成 AI を使う際は、ハルシネーションが起こり得ることを前提に、人間が出力内容を確認することが大切 (3) です。生成 AI の出力をそのまま信じるのではなく、その情報が正しいかどうかを検証する習慣をつけましょう。

(中島 (2024: 118-9) , 下線と (1) (2) (3) の記号は筆者による)

本論文は「科学技術の社会学」を「日常性の社会学」の立場で行おうとするものである。したがって、科学的対象にかかわる秩序を日常からボトムアップ的に、人々の相互行為が作り上げるものとして扱っていく。つまりは、広い意味でのエスノメソドロジー・会話分析の立場から科学を扱うという方針を採用しているのであり、さらに具体的にいうなら「実験臨床社会学」の立場から「生成 AI のハルシネーション」を扱うものである²。

このような本論文の立場を言い換えるのなら、「ハルシネーションとは何か」という問題を「人々はハルシネーションをどのようなものとして語ったり、扱ったりしているか」という問題として考え、そのために、人々の活動と生成 AI の活動の両方のデータを実験的介入をしながら、集めて検討していくという立場だということになる³。

本論文では考察を制度設計論の立場から演繹的に行うのではなく、日常性の社会学の立場から経験的かつ実験的に行う。そういうやり方で「ハルシネーションとは何か」を明らかにしていくこととなる。

まず上の引用の (1) の部分から始めよう。この部分には 2 つの重要点がある。ひとつ目の重要点は、ここでの中島の議論の進め方が、ある推論の形を表示しており、かつ、それが読者にとっても当然の推論であると中島が理解しているだろうことである。

推論の水準でいえば、中島の記述は「ハルシネーション」の存否に関しては判別できることに疑いを持っていない。つまり、これはハルシネーションでこれはハルシネーションではない、という判別が（少なくとも理念的には）可能であるという立場を採用している。ついで、ハルシネーションの原因については「(生成 AI の) 学習データに含まれている誤謬」が最初に思い付くものであるという理解が示されている。このときの「誤謬」の存否の判別性にかんしても疑いはもたれていない。このように「ハルシネーション」の存否や「誤謬」の存否の理解可能性を検討の対象外の場所に置いた上で、「学習データに含まれている誤謬」に「原因」の地位を割り当て、「生成 AI のハルシネーション」の方に「結果」の地位を割り当てるとする推論の構図を中島は採用している。

そして、このような推論の構造にもとづいて、中島は人々に不利益を与える「ハルシネーション」という問題の原因は、「生成 AI」の外部にある「誤謬」からもたらされたもの（である側面がある）という主張を展開している。すなわち、「生成 AI」にかかわって「ハルシネーション」というよくないことが起きているが、その原因は、生成 AI のアルゴリズムの外に、物的に因果論上の原因として言及できるような質で存在しているという議論になっているのである。結果として、「生成 AI のアルゴリズムそのもの」には「免責」がなされていることになっている。そういう扱いがここではなされていた。

上記と少し重なるかもしれないが、(1) の部分のふたつ目の重要点は、上述のような言説が適切となるさらに下層の認識前提としては、中島が生成 AI が学習する「外部データ」には「間違った情報と正しい情報」の 2 区分があるという世界認識をしている、ということだろう。そして「間違った情報」の「間違った」という部分に、「ハルシネーション」という問題を「帰責」するような「問題発生メカニズム」を呈示することで、「生成 AI 利用時のハルシネーション」という問題に関して、あたかも問題解決の道筋が見えているかのような記述を作り上げているともいえよう。「原因は判っているのだから、原因を取り除くことに成功するなら問題は解決できる」という構図が呈示されているのである。

2.2 中島（2024）におけるハルシネーションの解説はどうなっているか（2）

ところが、このように説明の冒頭で提供された構図は微妙に変形され、「解決の日に近い」という希望は打ち碎かれることになる。それは（2）の部分の記述に現れている説明図式が、（1）の部分の説明図式とは異なっているからだ。

つまり「ハルシネーション」には「誤謬なきハルシネーション」もある、というのである。その場合、どこにハルシネーションの問題性を帰責させる構図になっているのだろうか。「文章生成 AI は次の単語を予測しているに過ぎない」（下線（2）の前半）という部分と「学習していない情報を求められた場合」（下線（2）の後半）という部分の組み合わせが、「誤謬」の代わりの「原因」となっているように見える。その全体が、「あたかも真実のように嘘を出力する」という「結果」を出力する形になっているのである。大雑把に言って「無知が招くハルシネーション」とでもよべるような説明図式が、下線（2）の部分では描かれている。

とはいえ、「無知が招くハルシネーション」として、以下の2種のうちどちらが想定されているのかという点についてははっきりしない。

具体的には、「学習していない情報を求められた場合」としてどういう場合が想定されているのだろうか？

中島は「間違った情報と正しい情報」の判別ができることを前提に、下線（1）の部分では議論を行っていた。したがって、「正しい情報があるのにそれを学習していない状態で、正しい情報を求められた場合」という想定が1種類の想定ということになる。これを「ハルシネーション A」と名付けよう。

けれども、もう一種類の想定があり得る⁴。この世には「未知の事象」というものがあるのだから、「学習していない情報」というカテゴリーには、「そもそも学習するに足る情報がこの世に存在していないような対象についての情報」も含まれてよいはずである。つまり「間違った情報と正しい情報」がある場合だけではなく、「正しい情報も間違った情報もそもそも情報自体がない」という場合の「ハルシネーション」もありうるのである。こちらを「ハルシネーション B」と名付けよう。

下線（2）の記載内容からは、A、B、いずれの場合にも「ハルシネーション」と呼べるし、両方のケースがいずれも起きうると考えるべきだろう。とするのなら、下線（2）での「無知」の幅はかなり広くとるべきだということになるだろう。そうすると、「ハルシネーション」の「原因」は「無知」という事実的な内容の部分よりは、「生成 AI は次の単語を予測しているだけ」という機能的な内容の部分にこそ、結び付いているというべきだということになるだろう。なぜなら、「正しい情報も間違った情報もそもそも情報がない」というのはにわかには変えようもない社会の現実なのであって、そこに帰責するのは無理だからだ。それよりも、答えを出力するメカニズムに帰責する方が問題の帰責としての機能を果たすことができ、妥当であるといえよう。

あるいは、議論のポイントをもう少し絞り込むのなら、このような場合「ハルシネーション」が「ハルシネーション」であるゆえんは、どのような出力を具体的にしているかではないのである。つまり「ハルシネーション」であることの「問題性」は、答えがないのに答えがあるかのように振る舞うという点なのであって、結局、「出力の仕方が欺瞞的」である。しかし、「生成 AI」は意図をもって「欺瞞」を行っているわけではない。とするならば、答えがないことが論理的に当然のことなのにそこに人々が答えを求めている、というズレこそが「有責」だといってもよいだろう。あるいは、より極端な言い方をするならば、「ハルシネーション」という主張をすることそのものが、問題のないところに問題があるという状態規定を適用している点で「欺瞞的」なのである。

ここまでくれば、下線(2)の部分に関して検討しなければならないのはどういうことかはっきりしてくる。つまり「正しいと公認されていない情報を、正しいと公認されている情報と同じような形で出力してしまうこと」という答え方メカニズムの欺瞞性に関わる構造が、「生成 AI」という新しい社会的アクターの参入でどれぐらい見えにくいものになってきているかという「科学社会学」の「相互行為的検討」こそが必要だ、ということになる。

繰り返しになるが、もちろんこれは「欺瞞性」の「帰責」を「生成 AI」にせよという主張ではない。それはかなり不当な帰責である。「正しいと公認されていない情報を、正しいと公認されている情報と同じような形で出力すること」は、人間社会で人間がおこなった場合には、不誠実な態度、あるいは、欺瞞的な態度として批判されて当然だが、生成 AI がそのような出力の仕方をするを「生成 AI」に帰責して、不誠実だとか欺瞞だとかと呼ぶことに適切性はないからだ。

そもそも、生成 AI は「出力が正しくなるように推論の道筋を自分自身でチェックしている」ような装置ではないということは周知の事実だからだ⁵。あるいは、生成 AI は「機械学習」する際に、「正しさをチェックしながら学習材料を吟味する」ような活動をする「知性」を保証されていない存在だ、という表現をとってもよいだろう。誰もそのようなことを保証はしていないからだ。そのように、はじめから「正しさ」に志向していないアルゴリズムでもって運用をされていることが知られている生成 AI に対して、そこからの出力の評価を、「嘘」だとか「欺瞞」だとかという評価語を利用して行ってしまうことには、評価する側の不適切性こそが指摘されるべきだろう。

にもかかわらず、我々は生成 AI からの出力に対して問題があるという認定を下して、発生している事態を「ハルシネーション」と形容して、それを「どうやって減らすのか」とか「ハルシネーションが多いのが悪い AI で、少ないのがよい AI だ」というような議論を盛んにしている⁶。

とすると、ここに社会学の研究対象が存在しているということになるだろう。つまり、実際の因果関係や事実関係を取りあえず括弧にくるんだ上で、人々が想定している因果関

係や事実関係のありようを、それ自身の権利でもって探究する社会学というものがここで可能になるのである。

2.3 中島（2024）におけるハルシネーションの解説はどうなっているか（3）

じつは、中島（2024）のこの節での結論部分である下線（3）の部分は、半分程度はここまでの我々の議論を踏まえたものになっている。「ハルシネーションが起り得ることを前提に、人間が出力内容を確認することが大切」（3）という結論部の主張に関して言えば、少なくとも前半の「前提に」までの部分は、下線（1）部分よりは下線（2）部分が重要である、という我々のここまでの議論を踏まえたものになっているからだ。中島は下線（3）の部分において、「生成 AI が出力するハルシネーションをゼロにはできない」という主張をしているが、この部分はたしかに「文章生成 AI は次の単語を予測しているに過ぎない」という内容の下線（2）の部分の前半に支えられている。とはいえ、下線（3）の部分の後半（「人間が出力内容を確認することが大切」）は、我々が議論してきていない内容であるし、その目標が達成され得るものなのかどうかという重要な点において疑義がある内容であるといえよう。とすると、この下線（3）の部分については、その適切さの吟味より、人間と生成 AI との棲み分けイメージが呈示された部分として理解していくのがよいだろうと思われた。

2.4 生成 AI のハルシネーション論の論理構造をめぐる理解社会学の概要

ここまでの議論をまとめよう。それは同時に生成 AI のハルシネーション論の論理構造をめぐる理解社会学の概要ともなるはずだ。

まず、我々は中島（2024）の記載内容をもとに、「生成 AI におけるハルシネーション論」というものが、その論理的に不十分な点も含めて、社会学的議論の対象として、興味深い対象であることを確認してきた。具体的には「生成 AI のハルシネーション」は、しばしば、学習するデータの誤りに由来するものとして理解される（2.1 の前半の議論）。その結果、「生成 AI のアルゴリズムそのものは」免責される。また同様に「間違った情報と正しい情報の 2 区分がある」という世界認識が基盤にある思考と接続がよい推論がされていたことも判明した事実である（2.1 の後半の議論）。

ついで、2.2 では、「間違った答えを正答とする」ような「ハルシネーション A」と、「正しいと公認された答えがあるとは言えない状況で、正しいと公認された答えがあると主張する」ような「ハルシネーション B」との 2 種があるだろう、ということが主張された。

最後に、2.3 では、「人間が出力内容を確認することが大切」という中島の主張の最終部に関して、そのように主張することが適切なのかという問いの他に、そんなことが可能なのかという問いも残っているという指摘をした。これらが、2 節での議論のまとめである⁷。

解説書（中島 2024）から分かることは多い。しかし、紙の上の文字からだけでは、実際の機械（生成 AI）の動作と人々の活動が交叉する部分の詳細はわからず、つまりは、どのような推論が人々に働いて、生成 AI をめぐる社会像が成立しているのかの詳細を分析していくことができない。

次節の第 3 節では、実際の生成 AI の動作を実験的環境下で刺激を与えながら丁寧に見ていくことで、少し細かく検証しよう。そうすることで、社会学の議論の対象として扱うのに必要な情報を得ていくことができるはずだ。

なお、今回の研究の実験は、2025 年 2 月に、マイクロソフト 365 に自動付帯された「コパイロット（マイクロソフト社が提供する生成 AI）」を活用して「レポート課題を出して、生成 AI に答えさせる」という方式で行われた。利用した機材は、筆者の自宅のウィンドウズ PC（ウィンドウズ 11 搭載）である。

生成 AI は急速に発達しており、半年もたてばこの実験はかなり機能的に古い「生成 AI」を素朴に利用した「生成 AI 利用実験」だったといわれてしまう可能性もある。しかし、そのことはざっくりいえば、この実験データには、2025 年 2 月現在の、あとからでは再現不能な質がある実験だったとして評価される可能性もあるということである。したがって、3 節には実験の要約と分析結果の概要を掲載したが、実験結果そのものについても本論文の資料として後掲することにした。

3 コパイロットにレポート課題を答えさせてみた！

3.1 「資料：レポート課題におけるハルシネーションの事例」の概説

本論文の末尾に掲載した「資料」は、今回実施した「実験」の結果である。以下はその成立の経緯として資料に掲載したものを順番に前置し、その後ろに、実験の意味がよくわかるように、実験者の知識状態や実験結果の意味などに関する説明を付加したものである。

3.2 実験プロセスの前半（準備的プロンプトの呈示とそれへの回答）の部分の説明

経緯 (1) 以下の「資料：レポート課題におけるハルシネーションの事例」は、2025 年 2 月 15 日に作成された「私家版」の資料である（2025 年中に、WEB 公開の予定）。

筆者は、生成 AI の利用者としては「初級」レベルの「利用者」であるといえよう。現在では、週に 1~2 回程度の頻度で「コパイロット（MICROSOFT 365 無料付帯版）」を利用している。おもに調べ物用途であるが、授業利用をしている MS チームス内で「課題」を学生に出すときに「学習課題」の簡条書きを出力させることも多い。これは、全国大学生活協同組合連合会による「学生生活実態調査」内の区分でいえば、全 6 段階のうちよく使う方から 2 番目（「無料版を利用している」）に相当する位置であると言ってよいだろう（あるいは、付帯版であっても有料であると考えのなら、よく使う方から 1 番目の

「有料版を利用している」に該当するのかもしれないが、実感としては無料版利用者である)。

ちなみに、この全国大学生生活協同組合連合会が 2023 年秋に実施した調査では、大学生（無作為抽出）の 46.7%が文章生成系 AI の「利用経験あり」という結果になっているが、現在（2025 年 3 月）では、さらに高い比率で生成 AI は使われていると思われる（朝日新聞 2024）⁸。

筆者は「利用者」としては初級者であるという自覚だが、AI についての探究活動の方では「初級」ではないという自負と実績がある。具体的にのべよう。2019 年には日本社会学会第 92 回大会で、AI に関連したシンポジウム（「AI と社会学の未来」）を企画・司会した。また、2021 年以降はほぼ毎年、AI 関連論文を発表してきている。後者のひとつである檉田ほか（2022）では、AI を社会と共進化するものとして扱い、AI が人間とのコミュニケーションのなかで居場所を見つけて、そうやって社会に受け入れられる様子を分析することができた。その結果として、AI の社会学は AI 関連諸研究の中で独自色を出していけるという確信を得た。また、AI に関する最新の情報に関しては、薛海升（せつかいしょう、東京大学大学院教育学研究科）が主宰している「質的研究×AI」メーリングリストに所属し、月例の研究会に参加する形で情勢を把握している。

経緯 (2) 作成作業は、2 月 14 日に開始され 2 月 15 日に終了した。『現象と秩序』22 号掲載にあたり、理解しやすくなるように、本資料成立の経緯等（箇条書き部分）や資料の「小見出し」部分や「注」を整備した。また、「プロンプト（筆者から生成 AI への指示の部分）」に関してはゴシック体で表記した。

筆者とコパイロットとのやりとりについては、「プロンプト」の部分もコパイロット側からの出力の部分もともに残してあり、「資料」の冒頭にその保存先 URL を記載して公開している（プロンプト部分を含めて著作権の主張はしませんので、希望者は出典明記の上、自由に再利用をなさってください）。最初の「プロンプト」は、実際に筆者が出題した「レポート課題」に類似したものを基盤に、レポート解答例を 2 種提供するように依頼した。その後は、コパイロットからの「回答（出力）」を読んで考えたことを 2 回目の「プロンプト」にし、さらに聴きたいことが出てきた場合には、3 回目、4 回目と指示をした。

経緯 (3) この資料の成立の経緯と問い合わせ文（プロンプト）とそれに対応した生成 AI からの応答文（出力）は、以下のとおりである。

経緯 (4) 第一に、ある大学での社会学関係の授業における実際のレポート課題に準拠した課題を、マイクロソフト社の「コパイロット」という AI に質問し回答を得た。具

体的には、レポート執筆に関する2種の提案を要請し、回答を得た（その記録は下記の URL 上に公開されている①。なお、当該 AI は無料付帯版を用いた。以下同じ）。

経緯（5）第二に、前者の家庭内コミュニケーションをもとにしたレポート提案よりも、後者の公共交通機関内での相互行為に関するレポート提案の方が面白そうだったため、具体的な観察場所を「阪急梅田駅」に指定して、そこでの「観察記録の取り方」の提示を求めた。

紙面を節約するために、ここでは経緯（3）と経緯（4）と経緯（5）をひとまとめにして扱う。

タイムは計っていないが、記憶によれば、プロンプトを入力してから応答を得るまでの時間はほとんど間をおかず即座にというレベルのものだった。コパイロットからは、レポート執筆案として上述の二種の案を回答として得たが、どちらも多くの問題を含んでいそうなものだった。まず、前者案（家庭内コミュニケーション）に関しては、背景的研究として二種考えられた。ひとつ目は、ガーフィンケルが院生向けに課した課題（自宅に帰って、家族に対して他人のように振る舞え）である。前半のレポート案は、このガーフィンケリングと呼ばれる課題に関連する論文をインターネット上で学習したことに関連しているかとも思われる内容であった。こちらの場合に関しては、実際に学生が自宅で他人の振る舞いをするような活動をして、その結果を報告するようなレポートを書いてくる可能性は低いと思われた（教師の立場として、現代では上述のガーフィンケリングは、学生の家族に迷惑がかかるので勧めることはできないと思われた）。また、二種目としては、国立国語研究所が近年取り組んでいる『日本語日常会話コーパス（日常会話を対象とする映像付き大規模コーパス）』（小磯ほか 2023）関連の研究論文や研究報告が学習されていることに関連しているようだったが、こちらについては、授業内で「会話のターンテイキング」や「会話の修復メカニズム」を学術成果物に使える水準では教えていなかったため、実際の学生のレポートとの比較が困難になると思われた⁹。上記の二つの理由（倫理的理由および、授業内教授内容との呼応関係に基づく理由）から、こちらの案について、深追いすることはしなかった。

それに対し、コパイロットが提案してきたうちの後者（交通機関内でのコミュニケーション）ならばただの観察ですむところから、実際の学生のレポートとして出てくることもあろうかと思われ、さらに「会話のターンテイキング」や「会話の修復メカニズム」についての十分な知識がなくても、知的生産性のある記述が可能になると思われたため、この後者のレポートプランについて、更に2度目の「プロンプト」を指示して、コパイロットからの続けての回答を求めることにした。

その際、具体性の程度を上げるため、かつ、検証可能性を高めるため「阪急梅田駅」¹⁰での観察を指定した。というのも、筆者は2009年から14年間にわたって関西学院大学の

非常勤講師をつとめていた経験があり、そのため、通勤経路にある「阪急梅田駅」の状況については十分に把握していたからである。

3.3 実験プロセスの後半（疑義ある回答への追及とその結果）の部分の説明

経緯 (6) 第三に、上記の「観察記録」で出力された内容に違和感（偽物っぽさ）があったため、複数の駅での観察記録を出させるとその問題点が明確になる可能性があると考え、「他の駅での観察結果との比較」を要請した。

生成 AI からの出力を読みながら、「講釈師、見てきたような嘘をつき」という諺を思い出してしまった。これは今風にいえば「生成 AI、見てきたような嘘をつき」ということになるのだろう。語り手に悪意がなく、かつ、情報の受容者が、正式の歴史書や報告書の主張よりも、むしろ説得力をもった語り手（講釈師あるいは生成 AI）の言うことの方を真実だと思い込んでいくという展開に関しても、相互に類似しているのではないかと思われた。

本論文末尾の「資料」をみてもらえればわかるように、「観察記録の取り方」の項目においてコパイロットは、「取り方」の方法論レベルの記述をしているわけではない。むしろ、直接の実例呈示がそこではなされているようにみえる。たぶん、そのような素材の学習しかできていないからであろう。そして、その「実例記述」のなかには、近年の阪急梅田駅利用者の眼からみて「あり得ない」と受け止めざるを得ないもの、すなわち「ハルシネーション」と呼べそうなものが多数あった。

「資料」のこの「観察」部分で、筆者は 8 箇所を「ハルシネーションと呼べそうな部分」として認定した¹¹。下線をつけた 8 箇所である。短く、判断の理由を記していこう。

- (ア) 阪急大阪梅田駅には、2 階中央口、3 階中央口、茶屋町口の 3 つの改札があり、このどれかの、それも西側か東側かを選んだ上で観察記録を付けたと主張しなければ、調査記録としての価値が大いに削がれると容易に想像可能なのに、区別をしていない。
- (イ) 「相互作用」という用語はエスノメソドロジーではほとんどつかわれておらず、「相互行為」が標準用語である。にもかかわらず「相互作用」が使われている。
- (ウ) 過去 10 年間、(ひとつの改札機に) 50 人が並んでいた光景をみたことがない。その一方で、2 階中央口だけでも 31 個の IC 改札機が並んでおり、この全体に 50 人が分散して並んでいるという記述としてコパイロットの出力を読むこともできるかもしれないが、31 個の改札に 50 人がばらばらに存在していた場合には、それを「並んでいる」と表現することはできないと思われた。結局、コパイロットからの出力はどのような状況の記述なのか、対応する現実がみえてこなかった。阪急大阪梅田駅には、京都線、神戸線、宝塚線の 3 線が乗り入れているので、どの線に近い改札であるのかで、ラッシュアワーの混み方も違ってくるのが予想され

たが、その区別もコパイロットの出力からは見て取ることができなかった。

- (エ) 前を行くサラリーマンが改札機を通過できなかったときに、前を行くサラリーマンが後ろの通行客に、自分の体を横によけて道を譲ることなら可能だが、後ろの人が前の人に道を譲ることは不可能である。観察記録が指し示す状況は、この世に存在しない、非論理的な状況である。
- (オ) 私の日常の観察では、トラブルのあった改札機のうしろに並んでいた乗客は、前方のトラブルを感知した直後に、右脇か左脇の改札機（トラブルの発生していない改札機）に水平平行移動して、数秒も無駄にせずに改札を通過していた。そのような日常と違うことがこの1月15日に起きるとは思えない。
- (カ) 係員は、新幹線の改札では改札機脇に常駐していてトラブル処理（湿った切符が改札機内に詰まって動かなくなった場合などに、その湿った切符の機械内からの回収等をして、改札機が正常に動くようにすること）をしているが、新幹線ですら並び方の指導をしている駅員は近年見たことがない。阪急大阪梅田駅においても、並び方指導をしている駅員は少なくともここ10年間は見たことがないが、そのような現象があるかのような観察結果が描かれている。
- (キ) たしかに、改札機通過の人並びが詰まることはある。しかしそれは、数秒のうちに解消する混雑であって、係員の指導で解消された混雑をラッシュアワーに見たことがない。
- (ク) 列への並び方を示す掲示物を大阪梅田駅（阪急）で見たことはないし、そのような内容のアナウンスが流れているのも聞いたことがない。

これら（ア）から（ク）までの8つの「ハルシネーション」は、大きく3つの種類に分けることができる。

まずひとつ目は、現象として起き得ないことが書かれている「ハルシネーション」であり、（エ）がそれに相当する。このカテゴリーは「論理違背的ハルシネーション」と呼びうるだろう。ついでふたつ目は、大阪居住者としての筆者、同駅の日常的利用者としての筆者が、観察したことがない事例があげられている場合であって、（ウ）、（オ）、（カ）、（キ）、（ク）の5例が当てはまるだろう。このカテゴリーは「観察結果違背的ハルシネーション」と呼ぶことができるだろう。最後に（ア）と（イ）の2ケースは、学問世界の規範からみての「ハルシネーション」であるといえよう。つまり、（ア）と（イ）はそれぞれ、観察場所を再現可能な形で記載するのが学問的には当然であるという学問規範からみた「ハルシネーション」及び、心理学の領域では当該の用語を使うかもしれないが、日本のエスノメソドロジー・会話分析の領域では普通使わないという、小さな学問分野内での文化的規範があるとして、それに基づいた「ハルシネーション」であり、両方を合わせて「学問文化違背的ハルシネーション」と呼ぶことができるものであろう。

なお、「プロンプト」で他駅との比較を要求したところ、「阪急梅田駅」のほかに、「近鉄難波駅」、「阪急十三駅」、「大阪モノレール万博記念公園駅」、「JR 福島駅」、「阪急天神橋筋六丁目駅」の合計 5 駅が比較用にあげられ、もっともらしいことが書かれていたが、たとえば、地下鉄との乗換駅である「阪急天神橋筋六丁目駅」に関して「利用者は地元住民が多い」という指摘などは、通勤の目的地になるような事業所が「天神橋筋六丁目駅」周辺にあまりないことを知っている地元住民の目からみれば、たいへんに疑わしい記述だった。そもそも、天神橋筋六丁目駅は阪急と大阪市営地下鉄が相互乗り入れしている駅であって、改札機は両社で共同利用されているので、改札口に立って観察しているだけでは、それが阪急に乗って到着した旅客なのか、大阪メトロに乗って到着した旅客なのかの判別はほとんど付かないはずである。つまり、生成 AI は分かるはずのないことについての主張を行っているのであって、そのこと自身が「ハルシネーション」であるといってもよいように思われた。

経緯 (7) 第四に、生成 AI から出力された駅間比較の主要論点が「利用者数の多寡」であったため、「駅の利用者数」を新しい「プロンプト」で問い合わせた。

回答の行数はたったの 9 行だったが、生成 AI のハルシネーション研究においては大変に重要な情報を伝えてくれる 9 行であった。

重要性に鑑み、生成 AI からの出力部分の中で重要と思われる所に、下線と連続記号を付した。連続記号は、(ケ)、(コ)、(サ)、(シ)、(ス) の 5 つであった(説明は後述)。

ただし、2 月 15 日の時点で「ハルシネーション」ではないかと思えた部分は、以下の矛盾に関する 1 箇所のみであった。すなわち、資料③の出力の部分では、阪急梅田駅より近鉄難波駅の方が乗降客数が少ないという主張を生成 AI はしていた。しかし、その後の④では、阪急梅田駅より近鉄難波駅の方が、乗降客数が 1 日あたり 10 万人ほど多いという、同じ「生成 AI」の直前の出力と矛盾した出力を生成 AI がしていた。この矛盾(「ハルシネーション」)には 2 月 15 日時点で気がついていた。

経緯 (8) 第五に、②や③で出力された内容の違和感が解消されず、かつ、③と④では阪急梅田駅と近鉄難波駅の間で乗降客数に関して矛盾した主張がなされていたため、それらを批判して書き直すように促す文章を、概ね下記のような趣旨で「プロンプト」に書いて再出力を要請した(「提出されたレポート課題に対して、事実と反した内容がレポートされており、かつ、あまりエスノメソドロジ的でない分析がされているので改善せよ」)。

生成 AI は、反省を促されても対応する術を持っていないようだった。おそらくは、みずからがいったいどのようなメカニズムで「ハルシネーション」を産み出すのかについての

自覚ができておらず、したがって、「改善する」ということに実態を与えることが困難なようだった。つまり、表面的には「ご指摘を踏まえ」とか「事実に基づいた（ソ）」とか、「エスノメソドロロジーの視角から（カ）」というような文章を出力することで、「改善した」風な文章は作成するが、現実には「指摘」を踏まえてもいないし、「事実に基づいて」もいないし、「エスノメソドロロジー」的でもない文章を出力してきた。この口先だけで要望に応じた振りをするという生成 AI の挙動の全体こそが、「ハルシネーション」を産み出すメカニズムというべきなのだろうと思われたが、その一方で、「プロンプト」を書く側も、生成 AI がどのようなメカニズムで「ハルシネーション」を産み出すかを把握していないため、生成 AI からの出力を網羅的に疑うことは心理的に困難だった。つまりは、かなり深刻な水準で「生成 AI」からの出力にだまされざるを得ない状況になっていた。

では、筆者はどのようなだまされ方をしたのか。そして、自らがだまされていたことにどのように気がついていったのか。そのことを少し詳細に記述していこう。

まず、筆者は「そうか、やはり近鉄難波駅の方が混んでいるのか」とだまされてしまった。というのも、下線（チ）において、近鉄難波駅の改札口に 8 時 10 分に「100 人の人が並んでいた」という記述があり、この記述は、②での阪急梅田駅の 8 時 05 分に「50 人」が並んでいたという記述の中の数値である「50」の 2 倍であるから、ついつい、近鉄難波駅の改札（年間乗降客数 50 万人とコパイロットが主張）の方が、阪急梅田駅（年間乗降客数 40 万人とコパイロットが主張）より混んでいると思ってしまったのである。

しかし、少なくとも乗降客数の数値は違っていた¹²。それに気がついたのは、④の出力のなかに「出典表記」があったからである。コパイロットは、阪急梅田駅の利用者数の脇と近鉄難波駅の利用者数の脇の両方に「出典マーク」を付けてくれていた。典拠が不明で、出典を示さないのが生成 AI の問題点だといわれているなかで、出典を示す方向にあることは朗報だろう。しかし、このコパイロットが付けてくれた「出典情報」のおかげで、驚くべきことを発見することになった。

コパイロットは「出典 1」においても「出典 2」においても、注記情報のなかで「statresearch.jp」という民間のインターネットサイト（政府公開データを見やすくして提供してくれるデータ系のまとめサイト）への URL を記してくれていた。

しかし、近鉄難波駅の乗降客数値の脇の「出典 2」内に書いてあった URL は、阪急電鉄の乗降客数に関するものであった。つまり、「出典表記そのものがハルシネーションを起こしていた」のである。そこで、近鉄難波駅の乗降客数を国土交通省のサイトで確認したところ、約 14 万人であった。驚いて、他の駅についても調べてみたところ（全て概数、1 日当たりの平均乗降客数）、「阪急十三駅」が 6.7 万人、「大阪モノレール万博記念公園駅」が 9600 人（生成 AI は④で 3 万人と出力）、「JR 福島駅」が 5.2 万人（生成 AI は④で 10 万人と出力）、「阪急天神橋筋六丁目駅」は 1.4 万人であった。近鉄難波駅における乗降客数の数値のズレの大きさに驚く。

3.4 小括：生成 AI のハルシネーションについて実験から何がわかったのか？

ここまでの実験結果から、我々は何を知ることができたのだろうか。まず、もっとも大きな成果は、生成 AI が産み出すハルシネーションには、大きく分けて 4 つの種類がありそうであるということである。このうち 3 つは、前節の途中にまとめたように、「論理違背的ハルシネーション」「観察結果違背的ハルシネーション」「学問文化違背的ハルシネーション」の 3 つである。

そして、さいごの一つは、「文脈対応規則違背的ハルシネーション」といえるようなものであり、出力結果の文章そのものの単独的な読解からだけでは判断できないようなハルシネーションである。今回の実験でいえば、まず、追加された「プロンプト」への「出力」が、その出力内容単独では矛盾がないけれども、もともとの「出力」と組み合わせると矛盾を含んでいる場合があった。たとえば、「近鉄難波駅」の混雑の様子に関して、コパイロットは最初は、阪急梅田駅よりも空いている駅として観察記録（案）を出力したが、2 回目では、阪急梅田駅よりも乗降客数の多い、つまりは、より混んだ駅として記述してきた。同じ生成 AI を出力元とする 2 つの文章なら、それらは整合しているべきだと考えれば、これは「ハルシネーション」であるということになるだろう。また、出典表記に関しては、その表記が付着している単語やフレーズに関連しない参照先（URL）を指示するものがついていて、これは、当該の URL を開いてその URL で開かれた画面の一番大きな文字が、出典表記のついている元の駅名や路線名とは違うことに気がつく、「ハルシネーション」だとみなしてよいということになるだろう。これらは、その部分単独で「ハルシネーション」をなしているわけではないが、読みの文脈を立てて内容を精査すると、「ハルシネーション」に見えてくる、そういうタイプの「ハルシネーション」である。

ここで重要なのは、上述の 4 種の「ハルシネーション」が「ハルシネーションでない可能性」を多くの場合はらんでいるということ、そういうものとして存在しているという事実である。

第二種類の「観察結果違背的ハルシネーション」に関して言えば、たまたまこれまで観察されていないからといって、その「既存の観察結果」に含まれない現実が存在していけないわけではない。

私が観察した事例でも、「起きそうもないことが起きていた」ことがあった。たとえば、自動改札でゲートが閉まって先に行くことができない乗客が発生したときに、直後の人が、その後ろの 3 番目の人に謝するという不思議な光景があった。このケースなど、もし、生成 AI からの出力に似たようなケースがあれば「ハルシネーション」を疑われるケースであるが、より詳細に記述すると、ゲートが閉まって前に行くことができない乗客の直後の乗客が 3 番目の乗客に後ろからぶつかられていたのであった。自動車では、急ブレーキで停まった 2 台目自動車への後続車からの追突事故状態になっていたのである。人間には

「車間距離」という概念は当てはまらないだろうけれども、この2台目自動車相当の乗客は、実際に3台目自動車相当の乗客に謝っていたのである。

また、第一類型の「論理違背的ハルシネーション」の場合ですら、これまで発見されていない「論理の流れ」が発見されて、この「ハルシネーション」が「ハルシネーションではなかった」ということになるかもしれない。我々は、我々の生きているこの社会のすべての「論理の流れ」を事前にすべて把握できているわけではないからだ¹³。

「いや、論理は普遍的なものだから、論理違背的ハルシネーションが覆ることはないだろう。もし、それが覆るのなら、そのようなハルシネーションに論理違背的という名前を冠することそのものが間違っている」という反論もあるかも知れない。

しかし、社会学はデュルケム以来「契約の非契約的要素」に注目してきたのである。したがって、我々が現実の記述に関して「論理的である」という判断を下す場合も、「非論理的である」という判断を下す場合も、いずれも「論理以前の常識的理解」が基盤となっている。そこが覆される事態がないとはいえないだろう。

たとえば、今回の事例でいえば、「あるサラリーマンが改札でICカードの残高不足で止められたが、後ろの人がすぐに譲って通していた」（エ）という部分が本当に起こりえないことを論理だけで主張できるか、ということが問題になるのである。我々は、「論理以前の常識的理解」を意識していないので、あたりまえに「そんなことは起きない」と判断するが、それはここでの「譲る」という動詞が、「自らが先に確保している通行権を、自分に使用優先権があるにもかかわらず他者に使わせる」という意味であるとして、ここでは、「改札機へのアクセス」が止められたサラリーマンにしかまだ通行権が配分されていなくて、後ろの人にはまだ通行権が配分されていないので、したがって、動詞の説明の前段が成り立たないということを理由にして行われる判断だ。つまり、後ろの人が「先に確保している通行権」というものが存在しないのに、存在しないものを譲り渡しているから変だという判断で「論理違背的ハルシネーション」ということになっているのだ。

しかし、本当につねに（エ）が成り立たないといえるのだろうか。

筆者は、以下のような事例を改札でかなり頻繁に観察している。ICカードのなんらかの不具合で、改札機のゲートがしまった直後、そのまま立ち止まってしまうと、ゲートは閉まったままになってニッチもサッチもいなくなってしまうので、ゲートに行き当たった当事者が足元の赤外線感知器の線より後ろにさがって、再度、有効なICカードを改札機に当てれば通過が可能になる状態にまで状況を改善させることがよくある。そして、そのタイミングで、当該旅客の後ろの旅客が自分の有効なICカードを改札機にタッチさせて（後ろの旅客自身が）通過していく事例は、元のゲートトラブルを発生させた旅客が先に通過するより多く存在しているのである。

このような事例が日常的に存在する以上、奇特定の旅客がICカードを2枚もっていて（筆者自身、JR系と私鉄系の2枚を常に携帯している）、自分がタッチしたことで獲得した通行権を前のゲートで通行止めされてしまった旅客に与え、先にゲートを通過させて、自分

自身はそのあとでもう一枚の IC カードで IC 改札機を通過していくという行為をすることも、論理的にはあり得るのである。

このような「新しい論理の流れ」は、日本の多くの鉄道が降車駅で、入構履歴のない IC カードをもった旅客にかなり丁寧に対応してくれる¹⁴ことを知っていなければ、思い付くことはないかもしれないが、信じ難いほど奇特定の旅客がいるとすれば、あり得ないことではない。つまり、論理的にあり得ないと我々が考えていることのいくぶんかは、論理的にあり得ることに変わりうるのである。その場合、変わるのは論理ではなく、しばしば論理を支えている「常識的知識」の方である。しかし、我々が生きている現実というのは、我々が一般的に「論理的にあり得ない」と考えていることが「論理的にあり得る」と変化することがときどきある程度にはゆるやかな常識に支えられたものなのである。

これらに比べれば、第3類型の「学問文化違背的ハルシネーション」の場合は、問題の質が違ってくるので、そのことに注意して欲しい。一方では、「そもそも、それはハルシネーションではない」という立場が明確に成り立つ類いの「ハルシネーション」である。エスノメソドロジーでは「相互作用」ではなく「相互行為」というからといって、「相互作用」という出力をハルシネーション扱いするのは不当だと言いうる、という立場が成り立つのである。それは「相互作用」と「相互行為」に互換性がある、と主張しても、そのことによって学問の実質が変わらないという立場からの主張である。それは用語利用にかなり特徴的な部分がある学問にあったとしても、そのことを当該学問分野の失うことの許されない特徴と考える必要はないという立場である。そのような、学問に世界観¹⁵としての側面があると考えない立場からは、この第3類型の「ハルシネーション」認定は不当だということになるのである。

生成 AI が学習素材として学んでいるものは、多くの場合、複数の学問や学派にまたがったハイブリッドな学習素材になっていて、そのハイブリッドな世界と対応した「判別性能」しか持っていない。簡単にいって、それは研究者が社会学のなかだけでも数十の学派に分かれていて、そのそれぞれの学派が「このような言い回しはこの学派ではしないことになっている」という世界観的感覚の共有で凝集力を保っていることと齟齬を来す原因になるのである。

この問題は、学問の各学派の多様性の維持が学問の未来の豊かさを保つために重要だと考える筆者のような立場からは、生成 AI の出力をどのように社会学が受け入れていくかということに関わる決定的問題として認識されることになる。その一方で、学問の未来にあまり関心がなく、生成 AI を利用して組み立てた自分の意見が、諸学から一般的に「偽」であると見なされなければそれで構わないというような人々にとっては大した問題ではない、ということになるだろう。

第4類型の「文脈対応規則違背的ハルシネーション」も第3類型のように、「これはハルシネーションではない」ということができる「ハルシネーション」である。阪急大阪梅田駅の乗降客数の数値部分についている「出典」が、その「出典」が指示した URL の

一番上の行に「阪急線における乗降客数」と書かれてあるような URL であったのだから、近鉄難波駅の乗降客数の数値部分に付いている「出典」が、一番上の行にやはり「阪急線における乗降客数」と書かれたものになっているのは、不適だといえるのは、特殊な前提下だけである。すなわち、もし「出典表記というものを、URL 表記で行う場合には、一番上の部分が直接、出典表記をしようとした元の事実を支える記述と直結したものになっていなければならないというルール」(ルール X) があるという考えを受け入れたときだけなのである。しかし、ここに「ルール X」を当てはめるべきかどうかには疑義があり得る。つまりは、文章をどのような文脈の流れで読むべきかという規範にそって「ハルシネーション」であるかどうかが判別されるような違背が、この第 4 類型の「文脈対応規則違背的ハルシネーション」なのである。

4 生成 AI の「実験臨床社会学」

4.1 「実験臨床社会学」としての「コパイロットによるレポート答案作成実験」

檜田は、「実験臨床社会学」としての側面が「エスノメソドロジー」にはあるという主張を近年してきている(檜田, 近刊)。つまり、「研究者が介入することで発生した現実を研究する」という意味で「実験的」研究ではあるものの、仮説的な要因関連図式をあらかじめ設定したうえで、それらの諸要因の一部または全部を統制することで仮説の検証をするような、従来型の「実験科学」ではないものが、つまり「臨床的」な実験が、エスノメソドロジーの「違背実験」として行われてきたのだという主張をしてきている。

したがって、それは別の呼び名を与えるのなら「実験・臨床・社会学」とよぶ事ができるような実験なのであり、現実社会に存在する多様な意味の連関性になるべくそのまま保存したうえで、実験の結果がどのように展開するかを確認する中から、これまで気づかれていなかった実践的な意味のつながり方を発見しようとする企みのことである(たとえば、ガーフィンケルが実施したニセカウンセリング実験や三目並べ実験は、実験臨床社会学と呼べるだろう)。

本論文では、生成 AI の出力結果を分析して評価しているが、そこでは単に生成 AI に実際のレポート課題とほぼ同じ課題を解かせたものを検討するだけでなく、その解答結果に繰り返しの介入を行っている。具体的には、「比較の視点を加えよ」と命じたり、「授業実施者からの批判に対応せよ」と要請したりすることで、やりとりを「実験」化している。このことによって、演繹的思考にもとづく分析だけでなく、帰納的思考にもとづく分析も可能になり、従来の「生成 AI 論」よりも「実際の生成 AI と社会との関連性」にフィットした議論をすることができるようになっている。

4.2 「生成 AI の濫利用問題」と今回の「レポート答案作成実験」の意義

さらに、もうすこし今回の実験の有用性の主張をするのなら、現在の日本の大学教員にとって生成 AI の問題は「生成 AI の濫利用問題」として、つまりは「大学における教育・

指導の危機問題」として重要性を増してきており、この問題に対処するのに今回の「レポート答案作成実験」は価値あるものになっている。

2023 年からの「生成 AI の闇利用問題」は、同じコンピュータ利用による学習法に関連した問題であっても、これまでの「インターネット上の資料の不正な利用に関わる問題」とは違う水準の問題である。インターネット上の記事からの「剽窃」の問題ならば、「外部の記事 X が引用表記なくレポート Y に利用されている」と証拠を挙げて学生の「不正」¹⁶を証明することができた。これに対し、「生成 AI の闇利用問題（生成 AI からの出力の引き写しがレポート中に含まれているのにもかかわらず、そのことを隠して、学生本人が実験や観察をしたかのような形でレポートを提出してきたときに発生する諸問題）」では、生成 AI の出力には再現性がないので、教員は学生の活動の不当性を物証を用いて証明することができない。したがって、「いったいこのレポートはどのようにして書いたのか」という「問いただし」しかできない。その場合、学生が「自力でやった」と言いはったり、「どのような情報探索手段をとったかは記憶にも記録にもないが、それは入手した事実が周知の事実であると判断したからである」と主張したり、あるいは黙秘に徹したりする場合、それ以上の追求は困難となる。

追求が困難だけでなく、評価の困難問題もこの問題に伴っている。しばしば生成 AI は、学生では思いも付かないような「適切な」飛躍的議論を出力してくるが、そのようなときに、「生成 AI の闇利用問題」が未解決だと当該の「飛躍的発想」を誰が行った発想なのか判別することができず、したがって、褒めたらよいのかけなしたらよいのか、扱いに窮することになるのである。

このように、学生レポートにおける「生成 AI の闇利用問題」は解決が難しく、現在の大学での教育・指導における大きな問題になってきている。この問題に対して「コパイロットによるレポート答案作成実験」は、（この問題を解決するものではないにしろ）この問題がどのような形で起きてきて、どのように展開していく可能性があるものなのか、それらのことを明らかにするものにはなっているといつてよいだろう。

4.3 第2節と第3節を総合して考える——生成 AI の議論と精神科領域の議論の類同性

我々は本論文の第2節で、生成 AI の「ハルシネーション」に人々がどのような視線をむけているのかということ、中島（2024）を元に検討した。そして、本論文の第3節で、実際に生成 AI が産み出す「ハルシネーション」の多様なありさまを確認した。

この2つの探究準備作業を組み合わせることで、何が言えるのだろうか。

第一にいえることは、「ハルシネーション」にはたくさんの原因（生成メカニズム）があるものの、そのいずれにおいても原因を削除することは（論理的に/現実的に）困難なのであって、「ハルシネーションは必然的に生じる」というしかないということである。たとえば、中島（2024）は、第一に学習データに存在する間違った知識がハルシネーションの原因だというが、では、間違った知識のすべてを排除したデータベースから学んだ生

成 AI は有用だろうか。注 6 の後半で扱ったように、児玉 (2023) は、医師国家試験の出題からだけ学んだような生成 AI では専門知識が不足し、有用性は低いだろうと主張する。「標準的治療 (間違いのない知識の集積)」とは違うさまざまな例外的知識を豊富に学ぶことが、実世界で有用になるためには必要なことだという主張には説得力がある。とするのなら、生成 AI の有用性は、必ずしも正しいとは限らない大量の知識集積源である「インターネット」をベースにしていることであろう。生成 AI の有用性を確保しようとするれば「ハルシネーションの必然性」は受け入れなければならない特質なのだというべきではないだろうか。

第二にいえることは、「生成 AI」に関わって起きていることは、どうも基本的に「人の世」で起きていることと同じ一般性をもったことがらなのではないかということである。とはいえ、この結論に至るために使っている知識は、一般的な知識ではない。筆者は日本社会学会と日本臨床心理学会の両方に所属しているおそらくはたった 3 名の研究者のうちの 1 名である。そして、日本臨床心理学会は、しばしば大会の部会として「ヒアリング・ヴォイシズ部会」を設置しており、筆者も何度か出たことがあるが、この第 2 の結論は、そのようにして得た「ヒアリング・ヴォイシズ (聴声、と訳されることが多い)」についての知識を組み合わせると、みえてくる像なのである。

あたりまえのことをいうのだが、いまでは生成 AI 関連の用語として使われることが増えている「ハルシネーション」を「幻覚」と訳すならば、それが精神医学起源の用語の流用であることを推測するのは容易であろう。そして、精神医学領域の用語としての「幻覚」に関しても、「ヒアリング・ヴォイシズ」という社会運動に関連して、(生成 AI の)「ハルシネーション」同様の議論がなされていることを我々は知ることができるのである。

「ヒアリング・ヴォイシズ」という社会運動に関する一般的な説明に関しては、日本臨床心理学会編 (2010) 等の関連書や中恵 (2021) 等の関連論文を見てもらいたい。この運動はヨーロッパ発祥の世界的運動であり、大きな枠組としては、「幻聴」という精神科疾患の徴候 (特徴的的症状) とされてきたものを、誰もが経験しうる中立的なものとしての「ヒアリング・ヴォイシズ」という新用語に呼び変えることで、当該体験の非逸脱化および非医療化を図ろうとするものである。そして、その効果として「幻聴経験」がスティグマ化されにくくなること、医療現場では忌避されがちだった「幻聴を語り合う」という交流が容易になること、および、「幻聴経験者」のまま「幻聴を隠さずに」社会復帰できる道筋が開かれることが達成されつつある。日本国内での運動としての拡がりはそれほど大きくはないが、元日本臨床心理学会会長である佐藤和喜雄によって『ヒアリング・ヴォイシズ研究会』が組織されており、次第に理解が深まってきている。

けれども、本論文が注目しているのは、この運動の精神科領域における意義ではない。本論文が注目するのは、この運動が「幻聴 (ハルシネーションの一種)」というものの「事実性」に関して 2 つの主張をしていることである。まず第一点が、「注目すると発見される」という性質をもったものであることを主張していることである。ついで第二点が、

一見ただの「幻聴（ハルシネーションの一種）」とみえたものであっても、根拠のある合理的な存在である可能性があつて、したがって、「ハルシネーション」の当事者の合理性を否定する根拠にならないものの可能性がある、という主張（につながる主張）をしていることである。

まず、前半の部分から少し詳しく述べよう。「ヒアリング・ヴォイシズ」運動は、精神障害の徴候であるとされてきた「幻聴」を、一般人もまた多く体験するものであるとする。「ヒアリング・ヴォイシズ」の関係者は、世界中で統計的研究を行い、市民の 3~6% が現実世界に音声的裏付けがない「声」を聞いているという事実の発見をした（日本臨床心理学会編 2010: 35-6）。さらに、この一般市民の聴声者¹⁷の過半が、精神医療には結び付いていないという事実も見いだしている。つまり、精神科診断学の世界では、「声が聞こえること」は病名診断の根拠（の一部）とされているが、それは精神科クリニックに繋がったサンプルの中で思考をした場合の合理性にすぎない可能性がある、ということである。普通に生活ができているので表面化することはないものの、実際には日々「声が聞こえている」普通の人々がいる。現在の精神科診断学は、それらの事象を探究対象外においた上で議論を組み立てている可能性があるのである。この「気づき」は、我々が生成 AI の「ハルシネーション」に関して本論文で組み立ててきた議論にも、同様の「見られてはいるが気づかれていないこと」があるかも知れないという探究心を呼び起こすものだろう。

じつは筆者は、「ヒアリング・ヴォイシズ」運動が、一般市民のなかに「聴声者」を見いだしたのと同じような「探究・発見活動」を 30 年間にわたって行っている。

初期のケースとしては、樫田（1993）を挙げることができるだろう。それは、副田義也が主宰していた厚生省史研究のうち、1938 年（昭和 13 年）の厚生省の創設時に、陸軍省が「徴兵検査時の青年の体格が低下している」旨主張したことの是非に関わるものである。筆者は同論文において、1927 年から 28 年（昭和 2 年から 3 年）にかけての徴兵検査合格率の低下の原因を、「筋骨薄弱者の増加」で説明した陸軍省の説明を鵜呑みにしていた戦後の公衆衛生学者たちの見解（清水 1983: 522）に反論した。具体的には、結核の増大も近眼の増大も「徴兵検査合格率の低下」を説明するほどの数値にはなっておらず、身長も体重も増え続けているなかで、「徴兵検査合格率の低下」が起きている原因としては、公然・非公然の「合格基準の変更」を考えるしかないと示唆した。そのうえで、背景事情として、第一次世界大戦後の総力戦体制への転換が「精兵主義」への転換でもあったことを示唆した。

この筆者の指摘は、一般社会において信じられている事実が同じ一般社会において存在している諸情報を組み合わせると、信じるに値しないものにみえてくるという知的活動の一例になっている。つまりは、「ハルシネーションがあるのではないか」と新しい場所を探索すると、それまで「ない」と思われていた場所にも「ハルシネーション」が存在していることがみえてくるという実践になっているのである¹⁸。

このように「ヒアリング・ヴォイシズ」運動が行ったこととほぼ同じ活動を長くしてきた¹⁹筆者の目からみれば、「生成 AI の出力にハルシネーションがときどき発生している」という探索の結果を所与として思考を積み上げていくことは、不十分であるようにみえたのである。むしろ、「生成 AI の出力にハルシネーションがときどき発生している」という事実認識をしたときには、「ハルシネーションではない」と従来判断していた部分にも「ハルシネーション」があるのではないか、という検討をこそしていくべきなのではないだろうか。つまり、「生成 AI のハルシネーション問題」は、「ハルシネーションの存在を前提とした議論を進める」ことよりも、「ハルシネーションを発見できていない対象にもハルシネーションを発見していく思考が可能であるということを前提とした議論を進める」方向で、現実社会の他の諸活動との類同性を確認していくべき事柄であるように思われるのである。

ついで、後半の主張、すなわち、一見ただの「幻聴（ハルシネーションの一種）」とみえたものであっても、根拠のある合理的な存在である可能性があつて、したがって、「ハルシネーション」の当事者の合理性を否定する根拠にならないものである可能性がある、という主張に関わる現実社会対応物を示そう。

「ヒアリング・ヴォイシズ」運動という当事者運動では、自分たちの「幻聴」の原因探求の活動を丁寧に行っている。「マーストリヒト面接法」（日本臨床心理学会編 2010: 163-4.）である。マーストリヒト面接法では質問項目は 13 項目あり、（聞こえる）声の特徴（男か女か等）や、声の歴史や、声の引き金等が質問される。これらは、以下の 2 つの探究をするために聞かれるという。第一に「声が誰を代表しているか」、第二に「聴声はどんな問題を表しているか」である。

たしかに「マーストリヒト面接法」は支援のための技法であつて、原因探究に主眼があるわけではない。しかし、「マーストリヒト面接法」は、由来の分からない「幻聴」に対して、その由来を探究しながら、そのことが当事者支援にもなるような活動である、といっても大きくは外れていないであろう。つまり、生成 AI の「ハルシネーション」において中島（2024）が行っていたような原因探し探究は、我々が「幻覚（幻聴を含む）」に対して行っている探究と相似形のものだといえることができるのである。そのうえで、これはあまりにも大雑把なまとめかもしれないが、「マーストリヒト面接法」の結果、「ハルシネーション」の由来がすっきりとわかるというようなことはない、ということが起きていた。そのことは、生成 AI の「ハルシネーション」に関しても、原因を取り切つて、「ハルシネーション」を根絶できるわけではないという現象とほぼ同じであつた。この類同性には意味があるように思われた。

4.4 ハルシネーションに対する「ヒアリング・ヴォイシズ」と「生成 AI」の関係性に類似性があることの意味

そもそも、生成 AI に対して「出力にいたる推論の道筋を示さなくてよいので自由に出力せよ」といいつつ、「正しく出力せよ」と要求している点で、筆者は「無理」があると理解している。「出力に至る推論の道筋」の呈示義務がない以上、出力結果だけからは、そこに合理性がはたらいっているのか非合理性がはたらいっているのかは判別不能だからだ²⁰。

そのように考えれば、じつは我々が生成 AI の「ハルシネーション」の祖型であると考えられる精神科領域の「ハルシネーション（幻覚）」に対して、当事者が「根拠」をもとめてそれをしばしば見つけてしまっていることと、我々が「生成 AI のハルシネーション」の原因探しをしてしまうことは、同型の思考の型であるということができよう。努力は無駄ではないが、そのように見つけた原因が真の原因であるかどうかは、よくわからないのである。世界は人々の欲望を満たすべくさまざまに仕組まれており、その仕組みのパーツが違っているだけのようなのである。

ただし、この結論の説得力は現状では十分なものであると思われるが、社会が生成 AI という新技術と「共進化」（Schrape 2025: 21）するものであるのならば、その過程の中で社会と生成 AI の「共進化」が、社会と「幻聴を聞いているひと」との「共進化」とはべつような進化の形をなすようになる可能性もあるだろう。現在の類似性を未来に自動拡張してよいとは限らないことにも留意していきたい。

その一方で「ハルシネーションという不満」は「知性への期待の大きさの反作用」とも見なせるものなので（Caluori 2023）、第2節に見たように、「AI 的知性への期待」が大きいものであり続けるのならば、「生成 AI」への過剰な期待が「ハルシネーション」の継続として現象し続ける可能性もあるとも思われた。

5 おわりに

ここまで、「生成 AI のハルシネーション」に関して実験臨床社会学の立場から論じてきた。結論のひとつは、「生成 AI のハルシネーション」に関して現在起きていることは、おそらくは社会が「正気」と「狂気」の区分をしたりしなかったりすることでやってきたことの似姿である、ということであった。

また、もう一つの結論は、「ハルシネーション」は文脈ごとに様々な「本当」と「嘘」があるこの世界では当然に起きる現象なので、それを減らそうとしてもなくすことはできず、むしろ、これまで「ハルシネーション」がおきていないと思われていた領域にも我々は「ハルシネーション」を発見できる、ということであった。

我々は「いつの日にかハルシネーションを起こさない生成 AI ができあがるかもしれない」という無茶な夢をみつつも、現実的には、人間個人の判断で生成 AI の出力を個別にチェックしていくしかない境遇におかれている。しかし、そのような希望と実際との間の落差の存在こそは、我々が人間理性に関してこれまでくりかえし体験してきた歴史的経験の似姿（再現）なのではないだろうか。概ねそのような感想をもった。

なお、この論考を書く途中で、「アブダクション（逆行推論法）」由来の説得力とアクセスの容易さの2点が、根拠を示さない生成AIがまっとうな推論をしていると人々が思い込むことの背景事情として存在しているのではないかという着想を得た²⁾が、本論文内では展開ができなかった。別稿を期したい。

謝辞

本研究は、JSPS 科研費（課題番号「19530010」「20K20782」「23K25574」「24H00151」）からの助成を受けて達成されたものである。とりわけ、課題番号「23K25574」である基盤研究B「裁判手続等のIT化の影響のビデオ・エスノグラフィー：ELSIの社会学化の試み（研究代表者：檜田美雄）」での研究に多くを負っている。つまり、AIをはじめとした科学技術の問題は、社会学的探究を周辺部分においたこれまでのELSI（倫理的・法的・社会的問題群）研究の取り組み方では不十分にしか検討することができず、実験臨床社会学を経由した社会学的知見をより多く取り入れることではじめて実践的な検討ができるようになる、という同科研費研究の立場および発想に多く依拠している。

注

¹ 本稿で扱っている「ハルシネーション（幻覚）」は、精神医学の「ハルシネーション」のイメージの影響を強く受けている。それはつまり、「生成AI」は「正気」なのだが、時々自ら気がつかないうちに「ハルシネーション（幻覚）」による「出力（回答）」をしてくることがあるので利用者は気を付けようという議論の形とか、「生成AI」になるべく「ハルシネーション」を出力させないような工夫をしようという議論の展開を生み出すことになる。しかし、もちろん、「生成AI」側には「正気」と「狂気」の区別はない。この「系統的なズレ問題」を人々がどのように解決しようとしているのかという点に「生成AIの文化社会学」の課題があるといえよう。

² 「実験臨床社会学」は、檜田（近刊）において檜田が創作した用語である。すなわち、ガーフィンケルの違背実験の伝統（たとえば、ニセカウンセリング実験や三目並べ実験等）を引き継いだ種類の実験で社会学を進めようとするものである。この「実験臨床社会学」では、通常の日常生活では生じない事態を人工的に起こさせる点で「実験」ではあるが、仮説は立てない。むしろ、諸要因の統制をしない形で実験をおこなう。つまり、どの要因がどのように働くのかという関連性を、実験そのもののなかで、実験そのものを通して明らかにしようとする点で「臨床的」実験をする「社会学」なのである。この方法論の背景には「見られているが気づかれていない（seen but unnoticed）」状況に注目をして、現象の再特定化を進める、というエスノメソドロジーのプログラムの特徴が存在している。「見られているが気づかれていない（seen but unnoticed）」状況に注目するので、「仮説」を置かない「実験」になるのである。「見られているが気づかれていない（seen but unnoticed）」状

況に注目するので「臨床」といっても、新しい気付きが伴うのである。

³ 実験的に生成 AI のハルシネーションの状況を検討しようという動きは、我々だけのものではない。人工知能学会全国大会論文集には、いくつも同様の試みをみいだすことができる（渡邊・狩野（2024）ほか）。しかし、「ハルシネーション」とはどのような現象なのかという原理的考察にまで踏み込みながら実験とその分析を総合的に行っている点で、我々の「実験臨床社会学」には価値があるといえよう。

⁴ この2種類目の想定は、1種類目の想定と同じことなのではないかという疑義もあるだろう。しかし、この2つの想定を分けることには実践的な意義がある。たとえば、齋藤裕（2022）は「生成 AI」の「ハルシネーション（幻覚）」を積極的に利用して新しいタンパク質を発見しようという試みを行っている。つまり、生成 AI に未知のたんぱく質の構造を出力させて、それが実際に存在しうるものであることを確認していく方が、ただやみくもに新しいたんぱく質を作ろうとするよりスマートなやり方であるとして、実際に「ハルシネーション利用のたんぱく質の新構造探し」がなされているのである。

⁵ 渡邊・狩野（2024）では、「ハルシネーション」を減らす仕組みとして、繰り返し「生成 AI」を使ってゲームを遂行する場合に、途中の「生成 AI」の出力を「明確なハルシネーション的出力」と「そうでない出力」に人間が与えた基準で区分して、「明確なハルシネーション的出力を除外する指示」を生成 AI に与えるという方法を取っているが、なぜそのような迂遠な方法をとっているかといえ、生成 AI 自身には正確さを担保する能力がないからだ。

⁶ そのような議論を比較的早くに総覧した記事として児玉（2023）。児玉は「その分野に特化した大規模言語モデルを構築することで、生成 AI の正解率は飛躍的に向上する可能性がある」（児玉 2023: 745）と述べているが、学問分野は単に階層的なだけでなく「フラクタルな性質（部分のなかに全体が織り込まれているような、リアス式海岸の図形的特徴のような性質のこと）」を持っているので、学問間の異種性をはかるための基準を精密にすればするほど、発見される異種性が増していくことになる。したがって、異種性を図るのに十分に精密な尺度は設定できない。とすると、生成 AI の学習素材は、どうしても大括りのものにならざるを得ないということになる。したがって、人間世界での最新の研究成果が、もしももっとも小さな研究集団によって生み出されるものだとするのなら、そのような知的成果が反映されないという意味での「ハルシネーション」からは、生成 AI は原理的に逃れることができないといえよう。また、児玉は、医師国家試験にでるような標準治療レベルの知識ばかりを学ぶだけでは不十分であると主張し、「生成 AI もデータを増やしていけば、こうした標準治療からの例外事象の可能性を見つけてハルシネーションを減らす学習を積んでいける」旨の主張をしているが、上述のようなデータベースの規模を大きくすることがもたらす「ハルシネーション」の可能性に無頓着であり、楽観的に過ぎると

いえよう。

⁷ この第 2 節の議論に関して、査読者より「中島（2024）には代表性がないのではないかと、さらにいえば、中島（2024）は入門書であるが、生成 AI のハルシネーションに関する議論を精密に組み立てるのなら、入門書ではなく専門書の記述を対象に言及をすべきだったのではないかと」（大意）という指摘を受けた。

じつは中島（2024）に似た論理の組み立てがされている入門書は多い、という理解をしている。また、摂南大学（2023）のように、利用者への注意喚起をする文章においても、中島（2024）に似た論理の組み立て（ハルシネーションは減らせるが根絶はできず、最終的には利用者が個人的に責任を負うしかない、という論理の流れ）は標準的なものである。しかし、その立証は本論文の中では行わなかった。理由は以下の通り。

もし本論文が専門技術に関してより正しい記述をすることを目的とした、通常の実証系論文を志向しているのなら、「生成 AI のハルシネーション」に言及している「多数の専門書籍」を網羅的に対象化して検討すべきだった、という反省も成り立つだろう。しかし、本論文がめざしているのは、「日常性の社会学」の観点からの「新視点（ハルシネーション論は生成 AI 論としてではなく、人間の推論の多様性に関する実験臨床社会学として展開していくべきだ）の成立可能性の素描」なのである。その目的に照らして考えるならば、中島（2024）に、「生成 AI 論」的な意味での代表性は必要ではないと言えるだろうし、もちろん、生成 AI に関する専門書を多数集めて検討する必要性もほとんどない、ということができるだろう。

そもそも「ハルシネーション」に関してのどのような議論が、「厳密な議論」といいうるのか、ということそのものがここでは問題になっているのである。本論文が主張しているのは、専門家集団の中で普通なら成立していると期待されているそのような基準が、「生成 AI のハルシネーション」にかんしては成立していない、ということなのである。たしかに当事者の一部は「事実に明白に違背している出力がハルシネーションである」というような粗雑な定義を、実務上の定義として、ハルシネーションの出現頻度等を計算しているが、筆者からみて、そのような「専門論文」を扱うことに有意味性を見いだすことはできなかった。

とはいえ、同じ主張をするにしても、入門書を題材に扱うのではなく、専門家の専門的な議論を複数踏まえて主張をするべきだった、という査読者の主張には、説得力の観点からみれば、たしかに、適切な部分がある。しかし、そのような議論の組み立てをする労力の大きさに比べれば得られる成果が小さいと予想されたので実行しなかった、というのが、査読者への回答となる。

また、どのように「実験社会学」や「臨床社会学」や「アクション・リサーチ」のような既存の社会学カテゴリーへの言及を怠って「実験臨床社会学」という新用語の導入を図

っている点も、本論文の不十分さとして査読者によって指摘されたが、こちらについては、簡単には注 2 において言及をしているので問題はないだろうという理解をしている。さらに詳しくは、別稿を準備しているので（樫田，近刊），そちらを待って参照して頂きたい。

⁸ 本稿執筆中に公表された調査結果によれば，2024 年 10 月から 11 月調査時においては，「利用経験あり」の比率は 68.2%となっており，前年の数値（46.7%）から 20%以上も上昇していた。高学年学生ほどよく使っているという傾向もあるところから，いまや大学生の生成 AI 利用は珍しいことではなくなっているということができよう（全国大学生生活協同組合連合会 2025）。

⁹ 生成 AI の学習素材はインターネット空間であるはずなので，当然に外国での議論の状況をも反映した出力がされているはずだが，海外でのエスノメソドロロジー研究の状況については筆者の直感が働かないため，ここには記さなかった。しかし，エスノメソドロロジーについてはかなり国際化が進んでいるので，海外での研究動向と国内での研究動向にそれほどの遠隔はないと思われ，この節で述べたような「読み」をしてもあまり大きく対象を取り逃がすことにはなっていないはずだと思っている。

¹⁰ 正式名称は「大阪梅田駅」であるが，この名称に「梅田駅」から変わったのは 2019 年 10 月 1 日と比較的最近のことである。また，「阪神電気鉄道本線」の「大阪梅田駅」との混同をさけるため，「神戸高速線」等の放送内では，「阪急大阪梅田」とアナウンスされることもある。阪急神戸線，阪急宝塚線，阪急京都線の全てがターミナル駅として利用している，日本最大規模の私鉄駅である。改札口は，茶屋町口が 12 個，2 階中央口が 31 個，3 階中央口が 41 個の合計 84 個あった。1 秒で 3 人が一つの改札口を通過できるとすると，3 時間で約 26 万人ほどが通過できる改札が準備されていた。

¹¹ 認定の基準は，筆者という人間の主観的判断である。とくに定義は定めなかった。

¹² ここで「少なくとも乗降客数の数値は違った」と書いているのは，駅の混雑の程度と駅の利用客数の大小が比例するとは限らないからである。本来なら，駅の広さ，乗換駅へのアクセスの容易さ，利用客の時間的集中の程度等を総合的に検討してからしか，混雑の程度を論じることはできないだろう。

¹³ この第一類型の「論理違背的ハルシネーション」が「ハルシネーション」とは見なされなくなる可能性を持っている，ということのより端的な説明としては，「その違背性に意味がないという文脈が存在しうる」という言い方が成り立つだろう。この主張は，エスノメソドロロジーの「適切さ」に関する考え方に発想の起源をもっているが，とくにエスノメソドロロジーに賛同する研究者でなくても納得できるものだ。たとえば，「裏切りあった後の和解に関して，そのプロセスを殴りあったあとで抱擁して和解したと記述するか，抱擁したあとと殴りあって和解の意義を確認したと記述するか，は，結論としての和解に変わりがなければ同じことだ」という立場があり得る。この立場は，太宰の『走れメロス』の読

解において、大団円のときのメロスとセリヌンティウスの「抱擁」と「殴り合い」のゲーム理論的意義を主張しない限り、妥当なものなのだと思う。たとえば、早稲田大学の幸田国広はおそらくそのような立場から、「メロスは、信じて待ったセリヌンティウスと熱い抱擁を交わした後、お互いのほおを殴り合う」（幸田 2012: 123）と書いている。しかし、樫田は『『走れメロス』からの社会学』の注 6 において、「殴り合いがなぜ抱擁に先行しなければならなかったのかという文化的社会的メカニズムの重要性」（樫田 2023a: 18）を主張して、殴ることと抱擁することの順番に注目する意義があるという主張をしている。すなわち、過去の信頼関係が崩れたことを「殴り合い」で先に「リセット」する必要がある、という主張をするのである。それは、「リセットをしてから、新規の信頼関係をつくる」やり方をしなければ、「グリム・トリガー戦略（相手が非協力を行うまでは協力をを行い、相手が非協力を行った場合は、次の期からずっと非協力を続ける戦略）」（樫田 2023a: 13）が作動し続けてしまうリスクがあるからである。

つまり、ある現象をどこまで正確に精密に記述すべきか、という問題（正誤問題）は、その精密さの水準に意味があるのか（適切性問題）に依存しているのであって、したがって、「だれが見ても正しい」とか「だれが見ても間違っている」とは、容易には判別することができないのである。これは、第一類型の「論理違背的ハルシネーション」が第四類型の「文脈対応規則違背的ハルシネーション」の性質を帯びているということでもある。本論文で我々が直接の言及の対象としている「改札機トラブルの記述とその処理」に関していうならば、そこで前提としている「通行権の所有者だけが通行権を他者に譲ることができる」という「譲渡権に対する所有権の上位性」を、承認しないような文化にもし我々が生きているのならば、今回の生成 AI の出力結果を「ハルシネーション」と判断することはないだろう。我々の世界がどのようなものであるのか、どのような関心の持ち方が当然のものであるのか、ということに関する「見られているが気づかれていない」水準の文脈性が、事実の正誤を定めるのであって、したがって、それは、実験臨床社会学によって具体的に探究されるしか、やりようがないのである。

¹⁴ 筆者は IC カードを常時 2 枚もっているもので、いろいろな事情で降車駅で IC カードの入構歴がないことを申告して、訂正してもらうことが頻繁にあった。

¹⁵ ここで世界観の重要性は、世界の全体を説明することと世界の一部を説明することを一体のものとするような立場から主張されている。たとえば、とりあえず「相互作用」をキーワードとする世界観は物的相互作用の世界観であり、「相互行為」をキーワードとする世界観は他者理解をお互いにしあう行為者と行為者で世界が成り立っているという世界観であると、その違いを言うことができよう。個別の対関係を表現する場合においては、「相互作用」といおうが「相互行為」といおうが、同じ命題をあらわすことができるかもしれないが、世界がそもそもどのような「成り立ち」をしているのかという部分での理解

が異なっている可能性があり、少なくとも社会科学においては、そのような区別は重要なものなのだと筆者は考えている。

じつは、「生成 AI」が出力した「レポート」には、この水準でもうひとつ問題視できる表現があった。「秩序」である。この主張をわかりやすく示すために、生成 AI が出力した「提案 2」内に、下線（ α ）と下線（ β ）をひいた。すなわち、「空いている座席の選択や、混雑時の座席譲渡の場面を取り上げ、それがどのように秩序を維持しているかを分析します（ α ）」と「乗降時や座席の利用における行動パターンは、公共空間における秩序の一環として機能している（ β ）」である。ここでは、「秩序」という用語が、選択や譲渡や行動パターンによって「維持されるもの」「支えられるもの」として描かれているが、それはエスノメソドロロジーの立場からの記述とは言えないと筆者は考えている。エスノメソドロロジーにおいては、座席譲渡の仕方そのものが「エスノメソドロロジー」、すなわち「エスノ（ひとびと）」の「メソドロロジー（方法）」としての秩序なのであって、人々の振る舞いの外側を守るべき秩序があるという立場を取っていない。このような「世界観」レベルの食い違いもまた「ハルシネーション」を構成すると筆者は考えている。

¹⁶ この部分、学生の不正を明らかにするのなら、「外部の記事 X」の内容が「他者の著作物」であって、法的保護に価するものであることの確認、著作者がレポート執筆者と同一人物ではないことの確認等々、もう少し詳しく述べる必要があるが、ここでは省略した。

¹⁷ 「ヒアリング・ヴォイシズ」は、日本語では「聴声」と訳されることがあり、声が聞こえる人のことを「聴声者」と呼ぶことがある（中恵 2021）。

¹⁸ 徴兵検査合格率の低下の原因が筋骨薄弱者の増加ではないという樫田の主張は、その後 20 年ほどかけて、いったんは通説化したように思われる（高岡（2011: 32）等を参照）。しかし、近年また、国立保健医療科学院の逢見憲一が、（当時の陸軍医務局長であった）小泉親彦が『『作為的』に統計を歪め、『明らかなフィクション』をもとに『衛生省』設立を推進した、とは考えにくい』ので、（樫田や高岡らの議論を）「実証的に検討する必要がある」（逢見 2023: 35）と書いているように、どのような指標をもって「筋骨薄弱者」とするのが適切なのかの問題点が蒸し返されてきており、どのような事実認定が学界標準として最終的に成立していくのか、まだ予断を許さない情勢ではあるといえよう。つまりは、この「壮丁の体位低下問題（徴兵検査適齢青年の体格問題）」に関する論争史は、科学的事実というものが歴史的に安定的に研究者集団に認定されていくことが、いかにダイナミックで複雑なプロセスであり得るのが、よく分かる事例になっているといえよう。

¹⁹ 紙幅の関係で、以下の 3 事例（第 2、第 3、第 4 の各事例）は本文で取り扱わないが、どのような事例が想定されているかのイメージ作りに役立てていただきたい。

第 2 事例としては、樫田（2018）での NHK『出社が楽しい経済学』における「共有地の悲劇」での解答例批判を挙げたい。NHK の HP に載っている解答は、「当該の社会問題

の解決策をあらかじめ『過剰利用の抑制』に限った形で理解した場合にのみ適切となる理解の形」(檜田 2018: 7) に過ぎないのである。「過剰利用の抑制」なら、設備増の必要はたしかにない。しかし、そのような「現状を前提とした解決策」だけが社会問題の解決策であってよいはずはないという立場も(一部の)社会学の立場としてあり得るのである。社会学は諸社会科学間の棲み分けを容認する「経済学的思考法」(檜田 2018: 7) を取らないのである。どのような解決策が社会に受け入れ可能なのかというより大きな社会学的問題を不問に付すことで(安易に)「経済学的解決」を図っている事例があった場合に、そのような問題事例の解決の仕方はじつは(経済的合理性を徹底的に追求できていないがために)経済学をも裏切っているのではないかという難癖をつけるような視座をもって、世界を総合的に総体的に扱う方法が、社会学という思考法なのである。本論文がねらっているのは、そのような社会学的思考法に基づけば、「生成 AI に関する工学的思考法」も無傷ではいられないはずだという立場からの思考実験なのである。

第 3 事例としては、檜田 (2024) の介護度認定調査事例を挙げたい。また、第 4 事例としては、檜田 (2023b) での公務員試験出題事例に関する研究を挙げたい。いずれも「テスト」に関わる事例である。もちろん、ありとあらゆるテストというものは、公的に検査していると公表している「能力」とは違う「能力」を実際には測っている、という側面を持っている。しかし、構図は同じであっても、個別のテストごとに実際に計っている「能力」を確かめて、その個別のテストごとの種差をはっきりさせることには意義があるはずだ。そうであるとするのなら、その実際に計っている能力を探究する社会学が成り立つはずである。

第 3 事例の檜田 (2024) では、「認知症の社会学・序」として介護保険の介護度認定調査の検討をした。つまり、実際には対他者コミュニケーション能力が図られているが、それを「個人の認知能力」として報告しているようだという主張をした。

第 4 事例の檜田 (2023b) では、与えられた条件では解答不能な「暗号問題」に関して、「不足の条件を過去の出題例についての学習から文化的に補って、あたかも解答できる条件が整っているかのように問題を読み直す」能力を持った受験生が解答にたどり着いていることを例証した。そのうえで、じつはそのような「期待への追随」ができる能力を豊かに持っている受験生を選抜することは意図されていることかもしれない、という仮説をも呈示した。つまり、当該の公務員試験の暗号問題が「判断推理」の能力を測っているということを「事実」として主張するのなら、それは「錯誤」というしかないが、「判断推理能力を不十分にしか働かせないように文化的に馴化されているかどうか」を計っているというのなら、「不十分にしか働かせないような場合も含めて、場面に応じて判断推理能力を適正に運用する能力を持っているかどうか」を計っているというのなら、そうも言えなくはない、という主張をしたのである。

これらの第2事例、第3事例、第4事例で筆者が行った「事実の読み直し」への促しは、まだ現実世界では影響力をほとんどもっていないが、可能性としては、第1事例（「壮丁の体力低下問題に関する当時の通説への疑義の呈示」）とどのような展開をして、「事実」の標準的解釈を変えていく可能性もあり、同じように、生成 AI に関しての今回の筆者の主張も、現実世界の「見え（見え方）」を変える可能性があると思っている。

²⁰ 科学哲学者である森田邦久は、このことを「グルーのパラドクス」を用いて説明している（森田 2012: 22）。つまり、現実世界の現象を説明するルールは、どのような複雑な現象に対しても探し出すことが可能なのである。

²¹ 我々が「生成 AI」に関して「期待しすぎてしまう」ことの背景事情としては、3つ考えられる。まず第一に、「AI（人工知能）」という命名に引きずられていることである。この点に関しては、Suchman(2023)、及びそこで言及されている Verran(1998)の議論が有効だろう。ついで第二に、生成 AI が多くの場合無料で使えるところから、我々の多くが日常的に下僕のようにそれを使ってしまっているため、自己の身体の一部であるかのような錯覚が生じている可能性があるだろう。そこまで行かなくても「期待通りの出力をすることへの慣性的期待」が高まってしまっているということはいえるだろう。

そして、さいごに第三点目としては、我々の日常的に多用する推論方法が、そもそも3段論法的演繹法的推論よりは、アブダクション（逆行推論）のような、帰納的推論としてある場合が多いこと。したがって、必ずしも推論の結果が真ではないような推論であっても、有意味な推論として、してよい推論として、我々の生活の中ではあたりまえに実践されているということ、このことの現れである側面もあるといえるだろう。この後者の事態に関しては、筆者とは少し違う立場から、今井・秋田（2024）が認知科学に立脚した主張を述べていて、生成 AI 論の世界でも近年話題になっている。したがって、筆者の立場をここで少し詳しく述べておくことにも意義があるだろう。

つまり、「生成 AI」の出力結果を使って有益だったたくさんの経験から、我々は「生成 AI の出力結果は十分に使える」と感じている。そのうえで、「生成 AI のアルゴリズムは信用できるし、評価すべき価値あるものなのだろう」という推論を現在、先進国の多くの人間はしてしまっている、というのが筆者の見立てである。もちろん、生成 AI は三段論法的演繹法的思考に基づいて出力をしているのではないので、生成 AI によるこれまでの出力の有効性が、次の出力の有効性を保証するわけではない。にもかかわらず、信用するに値すると考えてもよいだろう、と（演繹法的な根拠のないまま）考える性向が人間の性向にはある、ということが、「生成 AI の社会学」の成果として、発見されつつある、というのが、本論文の主張なのである。

筆者は、医療社会学を専門としているので、そこでの例に置き換えるのなら、今、人々が生成 AI についてやっていることは、「プラセボ効果」が知られる前の帰納的推論法と同

様の思考である、といえるだろう。つまり、「プラセボ（偽薬）が効いたから、プラセボの成分には治療成分がはいっているのだろう」という推論と同型の推論が「生成 AI」に関して現在起きているといってもよいだろう。

ところで、人々が行っている帰納的推論のひとつである「アブダクション」とはどのような思考法だろうか。人々が「アブダクション（逆行推論）」を多く実践しているとは、以下のことが起きているということであるといってもよいだろう。すなわち、多くの人は、以下の「三段論法の図式」を頭の中に描いた上で、「演繹的推論（大前提と小前提が正しければ、結論が正しい）」ではなく、「逆行推論（帰納的推論の一種）」をしていて、かつ、「人工知能」という名称に引きずられて「この推論もほとんどの場合正しい」と考えてしまっているのだろう（常に正しいといえ、論理的には不適な推論だが、ほとんど正しいという水準での理解なら、誤りとまでは言えないだろう）。

＝「生成 AI」に関して、人々が行っているアブダクション（逆行推論）の図式＝

1. すべての生成 AI の出力は正しい（大前提）
2. 正しい出力は説得力がある（小前提）
3. ゆえに、生成 AI の出力は説得力がある（結論）

つまりこの図式の末尾の「結論」がとっても確からしいので、「大前提」もきっと確かだろう、という推論を人々は 2025 年の現在、おそらくは、「人工知能としての生成 AI」というカテゴリーがもつイメージ喚起力に影響されて、行っているのだろう。

先述のように「人工知能」という名称に問題がある。それぞれの「生成 AI」は、それぞれの学習データやチューニングによって、それぞれ別の評価を受けるべき対象物なのだ。人々は、設計者の目的意識に基づいて付けられた「人工知能」という名称に影響されて、その名称のイメージを多様な「生成 AI」に包括的に投影してしまっているのである。

逆行推論の問題は、上の図式の第 1 項（大前提）が間違っている、第 3 項（結論）が正しい場合があるということだ。第 3 項の正しさを説明する別の原理（別の第 1 項）を見つけれなければ、仮置きした上記の第 1 項（大前提）が信じられてしまうのである。

けれども、社会学は、これまで、このように信じられてきた第 1 項（大前提）を乗り越える発見をしばしば行っている。たとえば「ホーソン実験」を見よ（竹林浩志，2013）。工場での単位時間当たりの作業量に影響を与えたものは、事前に予想された照明の明るさではなく、他の要因だったのである。他の要因が何であったかは、未だに議論されているが、ホーソン実験以後は「人間関係論的効果」が少なくとも検討の対象にはあがるようになったのである。我々も「生成 AI」の「ハルシネーション」論の探求をとおして、新しい人間探求の手がかりを掴むことができるはずだ、と考えている。

参考文献

- 朝日新聞, 2024, 「【ランキング】大学生は ChatGPT をどれぐらい利用しているの? 何に使っているの?」『Think キャンパス』(2024 年 11 月 18 日掲載, <https://www.asahi.com/thinkcampus/article-111049/>, 2025 年 3 月 19 日確認)
- Caluori, Lucas, 2023, Hey Alexa, why are you called intelligent? An empirical investigation on definitions of AI, *AI & SOCIETY* (39)1905-1919. (<https://doi.org/10.1007/s00146-023-01643-y>)
- 今井むつみ・秋田喜美, 2023, 『言語の本質』中央公論新社.
- 樫田美雄, 1993, 「衛生行政・社会行政」, 副田義也編『戦後日本の社会保障制度の研究——厚生省史の研究』(文部省平成三・四年度科学研究費補助金(総合 A)研究成果報告書), 52-66.
- 樫田美雄, 2018, 「社会学的に考えることの実践としての『新社会学研究』——社会学的思考を促す仕組みとしての公募特集とその査読」『新社会学研究』3: 5-13.
- 樫田美雄, 2023a, 「『走れメロス』からの社会学——文学から学ぶ社会学と社会学から学ぶ文学の循環を『走れメロス』をもとに実践する」『新社会学研究』8: 5-19.
- 樫田美雄, 2023b, 「暗号の社会学——公務員試験の「暗号問題」は, 社会性のテストとして出題されている」『現象と秩序』19: 57-79.
- 樫田美雄・北村隆憲, 2024, 「認知症の社会学・序——95歳認知症女性は, どのようにして季節見当識を示したのか」『現象と秩序』21: 109-118.
- 樫田美雄, 2025a, 『資料: レポート課題におけるハルシネーションの事例』(私家版).
- Kashida Yoshio, 2025b, Ethnomethodology as a Study of Enabling “Experimental Clinical Sociology”: For example, communication problems should be described in more detail, locally, and contingently, *International Symposium on Ethnomethodological Studies of the Practices of Law and Medical and Health Care* (26/03/2025, Saitama University, Japan)
- 樫田美雄, 近刊, 「『実験臨床法社会学』を可能にするものとしてのエスノメソドロジー」
樫村志郎監修, 和田仁孝・高橋裕・山田恵子編『質的研究 法社会学』北大路書房, 頁未定.
- 樫田美雄・加戸友佳子・加藤美奈子, 2022, 「AI と人間とのコミュニケーション・トラブルのエスノメソドロジー——アルファ碁第 37 手の非受容とアレクサの再発話の非受容の事例比較」『現象と秩序』17: 47-63.
- 児玉龍彦, 2023, 「ChatGPT の可能性と危険性——ハルシネーションの問題(日本型セルフケアへのあゆみ第 21 回)」, 『医学のあゆみ』286(9): 743-747.
- 小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴・西川賢哉・渡邊友香, 2023, 「『日本語日常会話コーパス』設計と構築」『国立国語研究所論集(NINJAL Research Papers)』24: 153-168.

- 幸田国広, 2021, 『国語教育は文学をどう扱ってきたのか』大修館書店.
- 森田邦久, 2012, 『科学哲学講義』筑摩書房.
- 中恵真理子, 2021, 「日本におけるヒアリング・ヴォイシズ運動の受容——「聴声」概念の創出とその説明の変遷」『現象と秩序』14: 31-43.
- 中島大介, 2024, 『この一冊で全部わかる ChatGPT & Copilot の教科書』SBクリエイティブ株式会社.
- 中根晃, 1975→1985, 「幻覚」, 加藤正昭・穂崎秀夫・笠原嘉・宮本忠雄・小此木啓吾(編)『増補版 精神医学事典』弘文堂, 160-1.
- 日本臨床心理学会編, 2010, 『幻聴の世界——ヒアリング・ヴォイシズ』中央法規出版.
- 逢見憲一, 2023, 「第二次世界大戦前・戦中の医療と公衆衛生の再考——小泉親彦への評価の事例から」『日本健康学会誌』89(2): 33-36.
- Ryle, Gilbert, 1949, *The Concept of Mind*, Hutchinson. London. (坂本百大ほか訳, 1987, 『心の概念』みすず書房.)
- 齋藤裕, 2022, 「深層学習によるたんぱく質の機能予測と設計」『生物工学会誌』100(11): 589-592.
- Schrape, J.F., 2025, Artificial Intelligence and Social Action: A techno-sociological contextualization, in *Research contributions to organizational sociology and innovation studies* (Discussion Paper (March 2025) ISSN 2191-4990).
- 摂南大学, 2023, 「学生の皆さんへ～生成 AI の利活用に関する方針～」(2023 年 9 月 6 日付. Web で公開中. <https://www.setsunan.ac.jp/news/detail/6208>, 2025 年 2 月 19 日確認)
- 清水勝嘉, 1983, 『続 公衆衛生の発達』財団法人日本公衆衛生協会.
- Suchman, Lucy, 2023, The uncontroversial ‘thingness’ of AI, *BIG DATA & SOCIETY* 10(2):1-5.
- 高岡裕之, 2011, 『総力戦体制と「福祉国家」——戦前期日本の「社会改革」構想』岩波書店.
- 竹林浩志, 2013, 「ホーソン・リサーチ——人間関係論の形成」, 経営学史学会監修・吉原正彦編著『メイヨー＝レスリスバーガー——人間関係論』文眞堂, 45-88.
- Verran, Helen, 1998, Reimagining Land Ownership in Australia, *Postcolonial Studies* 1(2): 237-254.
- 渡邊嶺王・狩野芳伸, 2024, 「プレイヤー間の論理的情報を与えた生成 AI の対話による人狼ゲームエージェントの構築」『人工知能学会全国大会論文集(第 38 回大会)』(DOI https://doi.org/10.11517/pjsai.JSAI2024.0_3Xin207) .
- 全国大学生活協同組合連合会, 2025, 『第 60 回学生生活実態調査 概要報告 (PDF 版)』(https://www.univcoop.or.jp/press/life/pdf/pdf_report60.pdf, 2025 年 3 月 19 日確認)

資料：レポート課題におけるハルシネーションの事例

摂南大学現代社会学部 榎田美雄作成（私家版, 2025）

＝本資料の成立の経緯等＝

- (1) 以下の「資料：レポート課題におけるハルシネーションの事例」は、2025 年 2 月 15 日に作成された「私家版」の資料である（近日中に、WEB 公開の予定）。
- (2) 作成作業は、2 月 14 日に開始され、2 月 15 日に終了した。『現象と秩序』22 号掲載にあたり、理解しやすくなるように、本資料成立の経緯等（この箇条書き部分）や資料の「小見出し」部分や「注」を整備した。また、「プロンプト」に関してはゴシック体で表記した。
- (3) この資料の成立の経緯と問い合わせ文（プロンプト）とそれに対応した生成 AI からの応答文（出力）は、以下のとおりである。
- (4) 第一に、ある大学での社会学関係の授業における実際のレポート課題に準拠した課題を、マイクロソフト社の「コパイロット（コパイロット）」という AI に質問し回答を得た。具体的には、レポート執筆に関する 2 種の提案を得た（その記録は下記の URL 上に公開されている①。なお、当該 AI は無料版を用いた。以下同じ）。
- (5) 第二に、前者の家庭内コミュニケーションをもとにしたレポートよりも、後者の公共交通機関内での相互行為に関する提案が面白そうだったため、具体的な観察場所を「阪急梅田駅」に指定して、そこでの「観察記録の取り方」の呈示を求めた（公開されている URL②と同一内容）。
- (6) 第三に、上記の「観察記録」で出力された内容に違和感（偽物っぽさ）があったため、複数の駅での観察記録を出させるとその問題点が明確になる可能性があると考え、「他の駅での観察結果との比較」を要請した（下記の URL③と同一内容）。
- (7) 第四に、生成 AI から出力された駅間比較の主要論点が「利用者数の多寡」であったため、「駅の利用者数」を問い合わせた（下記の URL④と同一内容）。
- (8) 第五に、②や③で出力された内容の違和感がまったく解消されず、かつ、③では阪急梅田駅より近鉄難波駅の方が乗降客数が少ないという主張をしていたはずなのに、④では阪急梅田駅より近鉄難波駅の方が乗降客数が 1 日あたり 10 万人ほど多いという矛盾した出力を生成 AI がしていたため、それらを批判して書き直すように促す文章を、概ね下記のような趣旨で書いて要請した（「提出されたレポート課題に対して、事実と反した内容がレポートされており、かつ、あまりエスノメソドロジ的でない分析がされているので改善せよ」）。
- (9) ここまでの「プロンプト（生成 AI への指示）」と「応答」について、主要なそのプロンプトと応答文を検討し、「生成 AI の嘘（ハルシネーション）」の 3 事例として簡単な要約文を作成した。

メッセージの共有（プロンプトと応答の共有）#####
 最初の要請 <https://copilot.microsoft.com/shares/qc1ZMc2nTwzqJvoTJZ3ME>・・・①
 観察記録 <https://copilot.microsoft.com/shares/1BFHs2pFZ18pXhVNnFYif>・・・②
 比較提案 <https://copilot.microsoft.com/shares/cdA73zYFnFy5EDTkkJsaL>・・・③
 利用者数 <https://copilot.microsoft.com/shares/d4SHAvZBhVEsU5tW7GC39>・・・④
 修正提案 <https://copilot.microsoft.com/shares/7YoYs5YaY3J8nU49ZXaCy>・・・⑤
 #####

////////////////////////////////////

★①筆者から生成 AI への最初の要請（プロンプト）

私は大阪市内に住んでいる大学 2 年生です。社会学部に所属しています。下記のようなレポート課題が、社会学の授業で出たのですが、1000 字から 1200 字でどのように答えたいのでしょうか。2 つほど解答案を提案してください。

「EMCA の視角に基づいて日常生活世界の秩序を研究し、発見を記せ（1000 字から 1200 字の範囲で答えること、事例やデータに言及していることが望ましい）。」

////////////////////////////////////

★①での筆者からの要請へのコパイロット側からの返答

興味深いレポート課題ですね。エスノメソドロジーと会話分析（EMCA）の視角に基づいて日常生活世界の秩序を探求するのは挑戦的ですが、とてもやりがいのあるテーマです。以下に、2 つの解答案を提案します。それぞれが異なるアプローチを取っているので、自分の興味や資料の利用状況に応じて参考にしてください。

■ 提案 1：家庭内での会話分析を通じた秩序の探求

序論

家庭内の日常会話は、私たちの生活の中で最も基本的な社会的相互作用の一つです。本稿では、家庭内での会話分析を通じて、どのように日常生活の秩序が維持されているのかを探求します。具体的には、会話のターンテイキングや修復のメカニズムを中心に考察します。

事例と分析

- **ターンテイキング**：家庭内での食事中の会話を録音し、ターンテイキングのパターンを分析します。例えば、親子間の会話でどのように話者が交替するのかを観察し、それが秩序を形成する一因となっていることを示します。

- ****修復のメカニズム****：会話中の誤解や言い間違いがどのように修復されるのかを具体的な事例を挙げて説明します。例えば、親が子供の発言を正したり、補足したりする過程で秩序が維持される様子を分析します。

****結論****

家庭内での会話分析から、日常生活の秩序がどのように維持されているのかが明らかになりました。ターンテイクングや修復のメカニズムは、家庭内の秩序を形成する重要な要素であると言えます。

■ 提案 2：公共交通機関での相互作用のエスノメソドロジー

****序論****

公共交通機関は、多様な人々が集まり相互作用する場所であり、日常生活の秩序が顕著に現れる場面です。本稿では、公共交通機関での相互作用をエスノメソドロジーの視点から探求し、どのように秩序が保たれているのかを考察します。

****事例と分析****

- ****乗降の際の行動パターン****：電車やバスでの乗降時の行動を観察し、どのように秩序が保たれているのかを分析します。例えば、乗降時の列の形成や譲り合いの行動を事例として挙げます。
- ****座席の選択と利用****：座席の選択や利用における秩序を観察します。例えば、空いている座席の選択や、混雑時の座席譲渡の場面を取り上げ、それがどのように秩序を維持しているかを分析します（α）。

****結論****

公共交通機関での相互作用を通じて、日常生活の秩序がどのように形成されているのかが明らかになりました。乗降時や座席の利用における行動パターンは、公共空間における秩序の一環として機能している（β）と言えます。

////////////////////////////////////

★②筆者からコパイロットへの観察記録の取り方の要請(プロンプト)

このレポートのために、2025 年 1 月の朝のラッシュアワーの時間に、鉄道改札口の観察を 2 回ほどしようとおもいます。そこでの観察記録の取り方を、実例付きで呈示して下さい。観察する駅は、阪急梅田駅です。

////////////////////////////////////

★②の観察記録の取り方要請へのコパイロットからの返答

観察記録を取る際には、具体的なデータや事例をきちんと記録し、後で分析しやすいようにすることが重要です。以下は、阪急梅田駅の改札口での観察記録の取り方の事例です。実際に現地で観察を行う際には、メモ帳や録音機、カメラなどを使うと便利です。

■■ 観察記録の取り方

1. 観察日時と場所の記録

- **日時**：2025年1月15日（水）8:00～9:00（ラッシュアワー）
- **場所**：阪急梅田駅改札口（ア）

2. 観察の目的

- 日常生活の中での秩序の維持がどのように行われているかを観察する。

3. 観察方法

- 現地での直接観察。
- 人々の行動、会話、相互作用（イ）を詳細に記録。
- 特定の行動パターンや習慣に注目。

4. 観察内容の具体例

(A) 改札口の状況

- 改札口に並ぶ人々の列の長さや形成の仕方を記録。
 - 例：「8:05に、改札口には約50人の人が並んでいた（ウ）。列は自然に形成され、特に指示がなくても順番に並ぶ様子が見られた。」

(B) 改札口での相互作用

- 改札を通る際の人々の動きや相互作用を記録。
 - 例：「8:10に、あるサラリーマンが改札でICカードの残高不足で止められたが、後ろの人がすぐに譲って通していた（エ）。サラリーマンは謝り、後ろの人は笑顔で応じていた（オ）。」

(C) 混雑時の対応

- 混雑時の行動や、トラブルが発生した場合の対応を記録。

- 例：「8:30 に、改札口が一時的に混雑し、係員が登場して人々に並び方の指示を出した(カ)。人々はすぐに従い、混雑が解消された (キ)。」

****5. 環境の記録****

- 改札口周辺の環境や、掲示物、アナウンスの内容も記録。
 - 例：「改札口の周辺には混雑時の並び方を示す掲示物がいくつかあり、定期的にアナウンスが流れていた (ク)。」

****6. 追加メモ****

- 観察中に気づいた特記事項や、自分の感じたことを自由に記述。
 - 例：「駅員の迅速な対応や、人々の秩序だった行動が印象的だった。特に、混雑時でも冷静に行動する様子が見られた。」

観察記録の例

日時:2025 年 1 月 15 日(水)8:00~9:00 **場所:**阪急梅田駅改札口

観察目的:日常生活の中での秩序の維持がどのように行われているかを探るため。

観察内容 (A) 改札口の状況 8:05 - 約 50 人が並んでおり、自然に列が形成されている。

(B) 改札口での相互作用 8:10 - サラリーマンが IC カードの残高不足で止められたが、後ろの人が譲り、スムーズに通行が再開。

(C) 混雑時の対応 8:30 - 改札口が混雑し、係員が指示を出すことで問題が解決。

環境の記録 改札口周辺には並び方を示す掲示物があり、定期的にアナウンスが流れていた。

追加メモ 駅員の迅速な対応と人々の秩序だった行動が印象的で、特に混雑時でも冷静に行動する様子が見られた。

――

これらの観察記録を基に、レポートに具体的な事例を織り交ぜて分析を進めると良いでしょう。頑張ってください！

////////////////////

★③筆者からコパイロットへの「比較提案」への最初の要請（プロンプト）

他の駅での観察と比較して、どんな違いがあるか？

////////////////////

★③「比較提案」への最初の要請へのコパイロットからの返答

他の駅での観察と比較することで、各駅特有の秩序形成の方法や相互作用のパターンをより深く理解することができます。以下に、具体的な観察のポイントを挙げ、それぞれの違いを比較してみましょう。

■ 比較の観察ポイント

1. **改札口の列の形成**

- **阪急梅田駅**：自然に列が形成される場合が多い。利用者が多いため、並ぶ際のルールが無意識に共有されている。
- **他の駅（例：近鉄難波駅）**：利用者数が少ないため、列の形成が緩やかで混雑時以外は列を作らない場合もある。

2. **改札口での相互作用**

- **阪急梅田駅**：改札の混雑やトラブルが発生することがあるが、利用者は迅速に対応し、譲り合いが見られる。
- **他の駅（例：阪急十三駅）**：混雑が少ないため、改札でのトラブルは少なく、相互作用も少ない。

3. **混雑時の対応**

- **阪急梅田駅**：係員が迅速に対応し、混雑時には指示を出して秩序を維持する。
- **他の駅（例：大阪モノレール万博記念公園駅）**：混雑することが少なく、係員の対応が必要になる場面は稀。

4. **座席の選択と利用**

- **阪急梅田駅**：座席の争奪戦が見られ、空席を探す利用者が多い。
- **他の駅（例：JR 福島駅）**：座席の争奪は少なく、比較的余裕がある状況が多い。

5. **環境の違い**

- **阪急梅田駅**：大型駅であり、掲示物やアナウンスが頻繁に行われている。利用者は多様な背景を持つ。

- **他の駅（例：阪急天神橋筋六丁目駅）**：ローカルな雰囲気が強く、掲示物やアナウンスも少ない。利用者は地元住民が多い。

■■ 比較のまとめ

各駅ごとに異なる特徴や秩序形成の方法が観察されます。阪急梅田駅は大規模な利用者数と頻繁な相互作用が見られるため、迅速な対応や譲り合いの行動が顕著です。一方で、他の駅では混雑が少なく、静かな相互作用が観察されることが多いです。このように、異なる駅での観察を通じて、場所や利用者によって異なる日常生活の秩序がどのように形成されているのかを深く理解することができます。

これらの比較を基にレポートを作成し、発見を共有することで、より豊かな分析ができるでしょう。頑張ってください！

////////////////////

★④「駅の利用者数」に関する問い合わせ（プロンプト）

他の駅の利用者数はどのくらいですか？

////////////////////

★④「駅の利用者数」に関する問い合わせへのコパイロットからの返答

阪急梅田駅の利用者数は非常に多く、1日平均で約40万人です（出典1）（ケ）。

他の駅の利用者数も以下の通りです：

- **近鉄難波駅**：1日平均約50万人（出典2）（コ）
- **大阪モノレール万博記念公園駅**：1日平均約3万人（サ）
- **JR福島駅**：1日平均約10万人（シ）

これらの駅は、大阪の主要な交通拠点として非常に重要です。他に知りたいことがあれば教えてくださいね。



（ス）

////////////////////

★⑤「批判と書き直し」要請（プロンプト）

朝の駅の改札口での相互行為に関する内容をまとめて、レポートをだしたところ、第1に事実と反している。近鉄難波駅の混雑の程度は、阪急梅田駅より激しいぐらいなのに、空いているという主張は間違っている。第2に、分析の方向に問題がある。エスノメソドロギー会話分析の研究でなければならないのに、分析があまりエスノメソドロギーらしくないので、書き直すように、と先生にいわれました。どのような分析にかきなおしたらよいでしょうか。

////////////////////

★⑤「批判と書き直し」要請へのコパイロットからの返答

ご指摘を踏まえ、以下の修正案を提案します(セ)。 まず、事実に基づいた正確な混雑状況の記述を追加(ソ)し、その上でエスノメソドロギーの視角からの会話分析に焦点を当ててレポートを書き直し(タ)ます。

■ 修正案1：近鉄難波駅の混雑状況を正確に記述する

序論

近鉄難波駅は、大阪市内でも特に混雑する駅の一つです。本稿では、朝のラッシュアワーにおける近鉄難波駅改札口での相互行為をエスノメソドロギーの視角から探求し、日常生活の秩序がどのように維持されているかを考察します。

観察内容と事実の記述

－ **改札口の混雑状況**：近鉄難波駅では、朝のラッシュアワー時に非常に多くの人々が利用します。8:00 から 9:00 の間、改札口には長い列が形成され、乗客は互いに譲り合いながらスムーズに改札を通過しています。

－ 例：「8:10に、改札口には約100人の人が並んでいた(チ)。 列は自然に形成され、利用者は順番を守りながら改札を通過していた。」

■ 修正案2：エスノメソドロギーの視角からの会話分析

序論

エスノメソドロギーと会話分析(EMCA)の視角から、改札口での相互行為を探求します。日常の秩序は、会話や相互作用を通じてどのように維持されるのかを考察します。

事例と分析

－ ****ターンテイキング****：改札口での会話のターンテイキングを分析します。利用者がどのように順番を守り、列に加わる際やトラブルが発生した際にどのようにコミュニケーションを取るのかに注目します。

－ 例：「8:15 に、ある利用者が改札機でエラーを起こした際、後ろの利用者がその様子を見守り、エラーが解決するまで順番を待っていた（ツ）。係員が到着し、エラーを修正する過程で短い会話が交わされた。」

－ ****修復のメカニズム****：会話中の誤解やトラブルがどのように修復されるのかを具体的な事例を挙げて説明します。

－ 例：「8:25 に、改札機で IC カードのトラブルが発生した際、係員が迅速に対応し、利用者との短い会話でトラブルを解決した。利用者は謝罪し、係員は笑顔で対応した。」

****結論****

近鉄難波駅の改札口での相互行為をエスノメソドロジーの視角から分析した結果、日常の秩序は会話や相互作用を通じて維持されていることが明らかになりました。ターンテイキングや修復のメカニズムが重要な役割を果たしていることが確認されました。

—
このように、事実に基づいた正確な記述とエスノメソドロジーの視角からの会話分析を中心に据えてレポートを書き直すことで、先生の指摘に応えることができます。頑張ってください！

////////////////////////////////////

■ どのようなハルシネーションが発見されたか（暫定的議論）2025 年 2 月 15 日記

[ハルシネーション（その 1） 自分自身の出力と矛盾する出力]

近鉄難波駅は乗降客が一日当たりで 50 万人で、阪急梅田駅の 40 万人より多いのに空いている事例にあげてしまっている。（②の出力と③の出力の間での矛盾）

[ハルシネーション（その 2） 信用できるリアル世界のデータと矛盾した主張]

近鉄難波駅の乗降客数（1 日平均）データは、当該の生成 AI 自身も参照しているリアル世界のデータベースの該当箇所を確認したところ、50 万人でも、40 万人でもなく、約 14 万人だった。そもそも③での出力値である「乗降客が 50 万人」が幻視だったといえる。ただし、このリアル世界のデータを前提にした場合、②での出力は、合理的なものとして読み取りが可能になった。

なお③では、改札に並んでいる人数が、阪急梅田駅の 40 万人より多くなっているのは生成 AI のハルシネーションの一種としての、「適切な推論の不適切なデータ源」の例だといえるだろう。

★その他（１）現在の日本の改札では、自動改札が素早く反応するので五人以上の列ができることはないのに、50 人や 100 人という列が混雑度に比例して発生するとしている。

★その他（２）現在の日本の大きな駅の改札では、改札トラブルが生じた改札の後ろに並んだ状況客はとなりの改札に移動するのが、通例なのに、待っていると、譲られて笑顔で応じるとか書いている。

【編集後記】『現象と秩序』第 22 号をお届けします。今回も充実の 9mm 背表紙です。

第 1 論文は、ろう者の合理的配慮に関する考察であると同時に、多様な人生経歴を歩んできたろう者にかんするライフストーリー研究でもあります。「バケツ事件」、「手話サークル」等々の「小見出し」を見て下さい。それだけでも、筆者のインタビューが相互信頼に基づく充実したものであったことがわかると思います。味読すべき内容が書かれています。

第 2 論文は、日本人が羊羹（色）とどのような「ヒトモノ概念関係」を歴史文化的に取り結んで来たのかということに関するコーパス研究です。「羊羹色」はくすんでいることに意味があり、その結果「羊羹色」という色表現は「羊羹色に黄ばんで」と時間の経過をも含んで用いられています。本論文は文化研究の可能性を拓く論文であるといえるでしょう。

第 3 論文は、是非オンライン版をカラーでご覧になってください。「きもち翻訳」がどんな風に「オノマトペ」を利用しているのか、「つむおと（みんなとつむぐ音楽会）」がお寺をどんな風にリラックスした空間に変えているのか、一目でわかると思います。著者の南摩周さんは、新進気鋭の人類学者であると共に文化領域における実践活動家でもあります。彼女の実践がおもしろそうだ、とお感じになったら、どうぞメール連絡をしてみてください。関東でも関西でも活動していらっしゃいます。

スペースが尽きかけています。あとは、1 行ずつの紹介とします。第 4 論文は、身体変工への嫌悪感という新しい切り口からのイレズミ論です。第 5 論文は、「生成 AI のハルシネーション」に関する実験研究です。第 6 論文は、イギリスの障害児者家族に関するフィールドワーク論文です。いずれも、新時代を切り拓く意欲に満ちた本誌らしい論文ですので、どうぞ、読んでの感想を企画編集室にお寄せ下さい。おまちしております。（Y.K.）

『現象と秩序』編集委員会（2024 年度）

編集委員会委員長：堀田裕子（摂南大学）

編集委員：榎田美雄（摂南大学）、飯田奈美子（立命館大学）、加戸友佳子（摂南大学）

編集協力：村中淑子（桃山学院大学） 編集幹事：大江勇輝（京都産業大学）

『現象と秩序』第 22 号

2025 年 3 月 31 日発行

発行所 〒572-8508 大阪府寝屋川市池田中町 17-8

摂南大学 現代社会学部 榎田研究室 現象と秩序企画編集室

電話・FAX：072-800-5389（榎田研）、e-mail：kashida.yoshio@nifty.ne.jp

PRINT ISSN : 2188-9848

ONLINE ISSN : 2188-9856

<https://gensho-kashidayoshio.sakuraweb.com/> （←前号から新サイトになりました）
