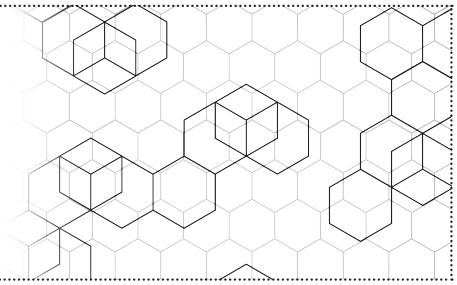
［連載］人間の言語能力とは何か――生成文法からの問い 番外編 ❷

大規模言語モデルは人間の言語能力の解明に役立つのか？（後編）

瀧川一学・折田奈甫



本連載第 2 回（2023 年 12 月号）の指定討論者である岡野原氏と瀧川氏は編訳者の一人である折田氏と，掲載した論考以外にメールでの議論も行った。前回に続き，番外編 2 としてその議論を収録する。

折田　連載第２回では，大規模言語モデル（LLM）を使って人間の言語能力について科学的な研究ができるか？ ということを議論しました。生成文法系の研究者が何より求めているのは「説明」ですが，機械学習モデルを用いて言語学者が納得するような，例えば言語獲得過程の説明ができるのでしょうか。

瀧川　機械学習それ自体は「予測」を与える技術であって「説明」や「理解」を与えようとするものではありません。科学的仮説は通常，観察事実によって検証されますが，機械学習は原理的に観察事実に合うように作られるため，ただデータを再現できるだけでは，現象を説明するモデルとは言えません。これは科学における伝統的なモデルとは大きく異なる特殊な性質で，十分な注意を払う必要があります。言語に限らず，化学者と話しても物理学者と話しても同様の議論になります。「説明」や「理解」は受け手である私たち人間の問題なんですよね。機械学習によって音声認識や画像認識はすでに実用レベルで実現されていますが，そのことが「私たちがどうやって音声や画像を認識しているか」の説明や理解に直結するわけではないんです。

折田　自然言語処理についてはどのようにお考えでしょうか。

瀧川　LLM 関連研究の隆盛が象徴するように，自然言語処理はずいぶん前から統計的機械学習と自然に融合しつつあると感じます。現在の LLM は，ホーンスティンの英語原稿を「日本語に直せ」とか「論点を要約して」とかだけではなく，「中学生にわかるように直せ」とか「5 文にまとめろ」とか「野球に喩えて説明して」とか「読書中毒で引きこもりのニートが斜め上から批判する感じで」とか「渋谷のギャルが友達に LINE する感じで」とか，無理難題も割とそつなくやってくれます。しかも，この指示文すら自然言語で自由に入力できて，LLM にとってはただの入力情報にすぎず，どこが質問や指示で，どこが本文かも明示的には与えなくてよい。このような柔軟で自然な言語運用が統計的な機械学習で実現できたことは大きな示唆だと思います。2012 年に辻井潤一先生が書かれた論考[1]で主な焦点になった言語の文脈依存性・多様性・制約大域性・長距離依存性などの問題は技術的には解消しつつあるのかもしれません。一方で，論考の最終節に挙げられた「プラトンの問題」や「記号と心的構築物との関係」などの難題は今も変わらず未解消で，論考冒頭のケネス・チャーチの主張のように「瑣末な改良の繰り返し」から離れて「経験主義以前の合理主義が取り扱おうとした難問を再吟味する必要がある」ようにも思います。

ピーター・ノーヴィグとノーム・チョムスキーの有名な論争が注目を集めていた 2014 年に，北海道大学で言語処理学会があって，辻井先生の招待講演を聴きに行ったのですが，「チョムスキーの“やり方”は間違っていると思うが，“プログラム／思想”はチョムスキーが正しいと思うんですよね」とおっしゃっていました。辻井先生は機械学習を言語に使う研究で著名で，ノーヴィグの肩をもつんだろうと思っていたので大変印象的でし

た[1]。

折田　辻井先生の学生だった岡野原さんはLLMが言語能力の解明に役立つというお考えをおもちです。一研究者として賛成はできませんが，私の知らない・見えていない世界をご存知の岡野原さんが強い信念をもってこうおっしゃることに敬意と関心をもっています。

瀧川　私も技術開発側なので，役立つ可能性には期待しています。「人と同様の言語処理は最終的には計算機で再現できると考えているし，人がどのようなアルゴリズムで言語を理解し，処理しているのかというのも全部明確になるのだと考えている」という展望には私は懐疑的なのですが，岡野原さん(やチョムスキー)のような強い信念こそが分野を駆動し発展させる原動力になっていることに深い敬意をもっています。先日，ドイツのゲッティンゲンに2週間滞在した際，宿泊していたホテルの横の広大な市民墓地に数学者ヒルベルトのお墓がありました。墓碑には有名な「Wir müssen wissen — wir werden wissen(我々は知らねばならない，我々は知るであろう)」だけが刻まれていました。この言葉は「科学には限界があり，いくつかの問題は永遠に解決できない(イグノラムス・イグノラビムス)」という懐疑論に対する有名な応答です。ヒルベルトの理想は弟子でもあったゲーデルの結果によって晩年打ち砕かれたと言われますが，ヒルベルト計画は20世紀の数学界を席巻し著しく発展させました。誰もいない森の中の墓に夢の跡のように残されたこの言葉を見て涙が出そうでした。

一方で，もし現在「AIですべて解ける」という謎の万能感や恐怖感を漠然と感じるとしたら，それは色めき立つビジネス界隈から発信される超デカイ声によって過剰増幅された風潮だという点を十分差し引いて考える必要があります。自分がよく目にする情報が全体を代表すると考えるのは典型的な認知バイアスです。「人工知能という分野が謙虚であったことなど一度もない」という批判を真摯に受け止め，安易な希望的思い込みを排し，長い間議論されてきた言語の困難さに慎重にかつ正確に向き合っていくことが健全だと思います。

折田　瀧川さんはLLMを使って自然言語の獲得・理解・使用などの問題を研究することについてどのようにお考えでしょうか。

瀧川　私は工学屋なので道具としての活用には肯定的です。ただし，人工知能分野に「あまりに擬人化が蔓延している」点は昨今非常に問題になっています*1。LLMを仮想的に人とみなし，今まで「人が使う言語」に対して培ってきた分析や評価を当てはめると，意味のない誤った解釈につながるので，機械学習という道具の特殊性を踏まえた慎重な利活用が必要だと思います。

折田　LLMのようなモデルが科学としての言語学や広くは認知科学で役立つとして，そこに理解や説明はあるのでしょうか？　説明がなければ科学ではないのでは？　でも，瀧川さんのご意見を聞いていると，説明を求める立場はどこか傲慢な気もしてきます……。

瀧川　「科学は理解や説明を求める」は本当にその通りです。現行の機械学習はそのままでは科学の目的に整合していないんですよね。「予測」のみで十分役割を果たしてきたビジネスや産業での活用を超えて科学研究にも使うならば，「科学の方法」の再考に加え，機械学習側でもまだまだ技術研究が必要です。

私の研究テーマは「記号や離散構造の機械学習」と「自然科学での利活用」で，工学に徹し，化学や分子科学を試験台として，人間が絡む問題(言語学や認知科学も含む)の研究を無意識に避けてきました。それで10年以上，様々な科学者と協働してきて痛感したのは，結局「科学も人間の営み」だということです。説明や理解を求めるのは私たちであって世界ではありません。科学とは何か，理解とは何か，説明とは何か，理論とは何か，を幾度も再考させられた10年でした。今回の言語の話も含め，多様な文脈でこうした問いに何度

＊1―https://www.technologyreview.jp/s/317295/large-language-models-arent-people-lets-stop-testing-them-like-they-were/

も遭遇し，人間が絡む問題を考えるスタートラインにようやく辿りついたのかもしれません。まだまだ道半ばです。

折田　言語学など人間が絡む問題は避けてきたとのことですが，機械学習がご専門の瀧川さんにとって自然言語とはどのようなものでしょうか？

瀧川　自然言語は，社会的に通じる範囲で自由に変形できる「雑で動的で泥臭い不確実なもの」と捉えてきたので，日常会話の大部分は統計原理でモデル化できてむしろ不思議はないなと思います。文法や語法も含む「記号使用の社会的取り決め」としての言語は，膨大な使用例から統計的に浮かびあがるのかもしれません。最近のLLMでこの感覚が半ば実証されたようにすら感じます。厳格な文法学者なら容認し難い変形・省略・逸脱は現実には日常茶飯事です。外国語での会話も文法学習より実際の会話で耳にする表現から始めるほうが近道ですよね。片言でも上手にコミュニケーションする人もいます。

その上でどうしても看過できないのが，言語の根底に「理路整然とした部分」がある点です。例えば，論理や数学やプログラミングの言語は，形式言語・人工言語として自然言語に対比されがちですが，改めて考えると自然言語がなければ存在し得ない人間の言語の産物です。デカルトが省察録で言うように，私たちは1000角形と1001角形の違いを経験できないし心像をもつことすらできませんが，その違いを概念的に理解することはできます。つまり，経験や心像に依存しない側面，統計原理に還元できない側面が言語の根底にあると感じるのです。純粋数学者ですら論文は自然言語混じりで書き，自動定理証明で得られた多数の定理が“良い”定理ではなかったことを考えると，非自然言語も人間の問題と無関係ではいられません。「説明」や「理解」は必ず言語で表出される必要があり，言語は私たちの理解や思考が成立しうるための必要条件であり限界なのです。

指定討論の原稿を書く際，後期ウィトゲンシュタインの『哲学探究』[2]と『確実性の問題』[3]を見直していました。原稿にも書いたように「意味を問うな，使用を見よ」と言い続けたウィトゲンシュタインの立場は「意味は全く扱わず」ただひたすら「言語使用の総体のみ」を見るという意味でLLMの統計原理と整合します。一方，何より重視された「生身の私たちが現実世界で織りなす実践」への接続は絶望的に欠けており，彼がLLMに関心を示したとは到底思えません。工学を学び論理学者を志した彼の関心が，癌と診断され死期を悟った最晩年に，統計原理で掴めそうな言語の「雑で動的で泥臭い社会的側面」ではなく，再び「理路整然とした確実性」へ向かっていく展開には本当に胸が熱くなります。『確実性の問題』は死の2日前まで書き続けた原稿です。私の研究上も興味は尽きません*2。

折田　LLMは「言語の理路整然とした部分」をもたずに学習し，表面的にはある程度自然に言語を使用しているように見えます。

瀧川　LLMと対話すると無数の人格の総体と話す感じがします。実際，仕組みはそれまでの文を見て違う人が一単語つけ加えていく連想ゲームのようです。毎単語違う人と会話していると思えば，文法は自然だけど，論理的一貫性がないとか，ハルシネーション／作話の問題とか，脱獄(本来は答えてはいけない質問に答えさせること)できてしまうとか，は不可避で当然に感じます。私たちの言動が論理的だとは全く思いませんが，「個」ゆえの実存的な惰性がある種の一貫性を生み出すように思います。統計は個を諦め消去する原理なので，この辺がなんだかLLMの応答が表面的に感じる遠因かもしれません。統計的な描像である大多数の平均とは「おもろない」凡庸さそのものでしょうから。

なお，LLMは文脈判断は得意なので，もし統計原理ではできない処理が必要なら検索や外部プログラムを適宜呼び出せば，実用上はあまり困らないように思います。こうした技術を入れるのもLangChainなどで簡単になりましたし，LLMの出力をLLMで評価したり，後段にファクトチェ

＊2―瀧川一学: 帰納と演繹の間を求めて: 記号と離散構造の統計的機械学習．電子情報通信学会コンピュテーション研究会，招待講演，2024年5月8日．https://youtu.be/T-grdDu3FIo

ッカーや検証器をつける研究も盛んです。

折田　LLMのような最先端の機械学習モデルであっても人間の言語や認知の問題をモデル化するのは難しい。この困難さって何なのでしょう？

瀧川　一つには，辻井先生が「“内的な処理”の計算モデル」[7]と呼ばれたように対象過程を直接観察できないからです。言語能力は心的状態，感覚，一人称性，指向性などとも不可分でしょうが，どれも直接観察できません。観察できる情報(言語使用の実例)だけから，モデル化のみならず，その是非の評価も行うしかありません。しかし，有限の事例の一般化は無数にありえますし(“一般化の問題”)，利用できる検証データも有限なので，その範囲では同程度に良いモデルも無数にありえます(“羅生門効果”“エピクロスの多説明原理”)。

もう一つには，それが開かれている系だからです。私たちは言葉を自由に組み立てて無限の表現を作り出せますし，言語自体も時間や環境で変化していくものです。ロケットは遠くの宇宙まで正確に飛ばせるのに，部屋のお片付けロボ激ムズ問題の困難と共通したものを感じます。起こりうる可能性が無限に開かれているのです。この点は「今わかっていないこと＝今データにないこと」をデータにもとづいて希求するAI for Scienceに通底する根本的な難しさだと思います。

折田　LLMで試せることはいろいろありそうで興味深いですが，科学的な問いであるのならば，先行研究をちゃんと踏まえて前提を明確にしてほしい，たとえばデビッド・マーの３レベルのうちどのレベルに相当するのかを明確にしてほしいとよく思います。計算理論レベルの人間の言語に関する問題で深層学習を使う根拠がわかりません。

瀧川　最初の話題の繰り返しになりますが，視覚に説明を与えようと考えたマーとは違い，今の機械学習は説明や理解を目的とするものではありません。LLMは機械学習なので，そのまま科学的モデルとも見なすのは，飛行機を鳥の飛行の「説明」と見なすのと同種の概念的混同です。統計原理を使えば，モデルが対象の現象を捉えていなくても「予測」が可能な場合はあるのです*3。現行の機械学習を牽引してきたのはITテックで，LLMもTransformerもそこ生まれです。科学者が考える以上に，高い精度の予測さえできれば十分な状況も多いのです。機械学習では，解釈と予測の非両立性と呼ばれますが，解釈性を諦めることで「予測」の精度は飛躍的に向上してきました。

また，現在の深層学習は神経回路のモデル化から離れ工学的観点で設計されるもので，マーの3レベルの1〜2層に対応すべきだとも一概に言えません。辻井先生の論考[7]はこうした言語学における経験主義vs.合理主義を概観していてマーの3レベルも取り上げています。経験主義に立てば，マーの合理主義(計算主義)の枠組に沿う必要はありませんが，敢えてそう解釈するにしても，3レベルとも共通の仕組みでいけるはずと考える人も多いように感じます。マーではなくシステム1／システム2のダニエル・カーネマンに耳を傾けているようです。

折田　解釈と予測が両立しないのは困りますね……。最近のニューラルネットで認知科学的研究をする研究者たちにとっては，マーの３レベルなんてどうでもいいのだろうなと思ってはいましたが腑に落ちました。

瀧川　もし予測精度の向上には複雑なモデルが必須で解釈性は損なわれてしまうのなら，たとえ真の法則が存在するにしても，それは私たちが理解できるほどシンプルではない可能性を示唆するようにも思います。学習済みのGPT-4は多数の項をもつとは言え，ただ一つの数式でそれは実際に既にそこにあるものです。その数式では理解にも説明にもならん(もっと簡潔に噛み砕け)というのは，単に私たち側の認知能力の問題かもしれません。

折田　人間の認知能力の問題はあるにしても，インプットの質と量が人間の子どものそれと乖離していたり，言葉の学習でいえば，人間の子どもの学習方法や学習順序とは明らかに異なるなど，最新のニューラルネットモデルを用いた研究であっても，言語獲得の研究者から見ればツッコミどこ

*3―https://videolectures.net/kdd2018_hand_data_science/

ろは多いです。人間が解釈できるようなモデルではないから説明にならないという問題以前に，これまでの研究の蓄積との不一致があります。しかし，私のような言語学者からの批判は取るに足らず，そんなことよりも経験的に予測精度を上げるということを突き詰めてから見える世界を目指しているのでしょう。

最後に，本連載第２回の指定討論には入りきらなかった重要な論点などあればお願いします。

瀧川　LLMを訓練データの情報圧縮だとみなす考え方が広くあります*4。圧縮としてみれば，明らかに非可逆(lossy)なので，JPEGの復元ノイズのようなアヤがハルシネーションみたいな現象ともみなせます*5。人類が生成してきた膨大な量のテキストの圧縮下限が，私たち個人の認知容量内に収まる保証はどこにもありません。私たちに理解可能な簡潔な「説明」があるはずという前提は，オッカムのカミソリ以来の希望的信心に過ぎないのかもしれません。

２点目は機械学習モデルのバイアスの話です。Transformerに内在する帰納バイアスは極めてマイルドなもので，デプロイされたLLMのバイアスの多くが開発に使用した訓練データ・検証データ由来のものだという点です。つまり，データを準備するときに「人間が入れたバイアス」だということです。人が集めるデータには必ず何らかのバイアスが含まれ，観察研究だけでは厳密な因果推論はそもそも不可能です。これを回避できる奥義(あらゆる交絡因子を無効化する奥義)は介入研究を伴うランダム化比較試験(RCT)だけです。が，機械学習に使う大規模なデータや，人を対象とする認知科学実験では観察研究にならざるを得ない制限があります。例えば，言語獲得の研究をRCTで実施するには，言語獲得に失敗する被験者を許容する必要があり，倫理的に不可能です。認知科学や心理学だけではなく，ヒトを対象とする医学や教育学なども同じ困難を共有しています。これは物理学を筆頭とするいわゆるハードサイエンスと大きく異なる点です。人間が絡む問題の研究では，取れないデータがほとんどであり，手に入る範囲の(偏った)データにもとづいて何かを立証するしかないという大きな制約があることを常に意識すべきです。

最後に，訓練データと検証データは常に有限なのに対して，システムが運用時に出会うデータには無限の可能性がある点です。機械学習モデルを実世界投入すると，この意味で「常に」過小決定(underdetermined)な設定になります。つまり，機械学習モデルは，背景に真の法則がある場合でも，開発時試験の範囲外ではそれを捉えてはいないと考えるほうがよいということです。これは科学や経済など実応用では昔から指摘されてきた問題*6でAI for Science研究の中心課題です。また，この問題は様々な実世界応用で顕在化していて*7，いわゆるAI Safety研究の中心課題でもあります。

文献

1―辻井潤一：合理主義と経験主義のはざまで．人工知能学会誌，**27**(3)，273(2012)

2―L. ウィトゲンシュタイン:『哲学探究』．鬼界彰夫訳，講談社(2020)

3―L. ウィトゲンシュタイン :「確実性の問題」.『ウィトゲンシュタイン全集(第９巻)』所収．黒田亘訳，大修館書店(1975)pp. 1-169

タイトル画像クレジット：vladystock/123RF

＊次回(９月号掲載予定)は番外編３として折田奈甫・次田瞬・窪田悠介の三氏による鼎談(前編)をお届けする予定です。

瀧川一学　たきがわ　いちがく
京都大学国際高等教育院特定教授，北海道大学化学反応創成研究拠点特任教授(機械学習・機械発見)

折田奈甫　おりた　なほ
早稲田大学理工学術院英語教育センター准教授
(第一言語獲得・心理言語学)

＊4―https://arxiv.org/abs/2309.10668

＊5―https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web

＊6―例えば，G. E. P. Box: J. Am. Stat. Assoc., **71**, 791(1976)やD. Hand: Harvard Data Science Review, **1**(1)(2019)など。

＊7―https://research.google/blog/how-underspecification-presents-challenges-for-machine-learning/