

# 国立国語研究所学術情報リポジトリ

## 【鼎談】言語学と機械学習の「共闘」は可能か？（後編）

メタデータ	言語: Japanese 出版者: 岩波書店 公開日: 2024-11-01 キーワード (Ja): キーワード (En): 作成者: 折田, 奈甫, 窪田, 悠介, 次田, 瞬 メールアドレス: 所属:
URL	<a href="https://repository.ninjal.ac.jp/records/2000327">https://repository.ninjal.ac.jp/records/2000327</a>

## 【鼎談】言語学と機械学習の 「共闘」は可能か? (後編)

折田奈甫・窪田悠介・次田 瞬

次田 そこで問題になるのが、連載第2回の瀧川さんの指定討論の末尾に出てくる「共闘」の提案ということですね。「生成言語モデルも、プログラミング言語も、論理と数学の言語も、私たちの言語活動の産物であると再認識すれば、この奇妙な呪文の体系もまた新しい言語の地平と言えるのではないだろうか。ウィトゲンシュタインは「私たちは言語と戦っている。私たちは言語と交戦中である。」と書き残している。だとすれば、私たちは共闘すべきである。」

折田 そうなんです。共闘についてお二人がどう思ったか聞きたかったです。

次田 共闘するにも勢力が違いすぎるのでは……。

折田 そうなんですよ！ 我々めっちゃお金もない、人もいない、ビッグな発見とかもこしばらくない、そういう言語学者と、第3次人工知能ブームでイケイケブイブイなNLP・機械学習は共闘できるんでしょうか？

次田 瀧川さんのメールに、「合理主義と経験主義のはざま」<sup>1</sup>という辻井潤一先生の論文への言及がありましたよね。少し目を通したんですけど、論文の序盤で「(1)経験主義の言語処理が主流になり過ぎた。(2)経験主義的な方法で成果があがる簡単な問題はすでに研究し尽くされた。(3)残された問題の解決には、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。(4)これらの難問への取組みを回避した(あるいは、その経験に無知な)経験主義からの研究は、瑣末な改良の繰返しになる」という論点整理がある。10年前にこういうことが言われたっていうのが、なんだかつらいですね。合理主義への揺り戻しはこなくて、むしろ10年前と比べて状況はさらに……。結局、合理主義に用はなかった

てことになってしまいそうです。

折田 つらい話です。この辻井先生の論文、まだ10年前なんですね！ もっと前だと思っていた。衝撃……。辻井先生の若い頃の話、聞いたことがあります？ 辻井先生、NAIST(奈良先端科学技術大学院大学)の自然言語処理の松本裕治先生と言語学の先生たちは、若い頃いろいろ活発にディスカッションされてたみたい。いい時代！

窪田 みんなで集まって、多分今我々がやっているような感じで好き放題言い合って、それで飲み屋に繰り出すっていうのをしょっちゅうやっていたみたいな話を聞いたことがある。

次田 つまり共闘していたのは昔だった、と。

窪田 辻井先生、昔からこういうこと言いつけてますよね。その意味では、辻井研ってNLPの世界の中でも立ち位置が特殊ですね。

折田 辻井先生の学生だった宮尾祐介先生は、いつかのトークで、生得的な言語知識は必要だと思いますとおっしゃっていた記憶があります。辻井先生のエッセンスみたいなのは、学生には引き継がれているようですね。でも……ねえ。今の時代、NLPerでニューラル使わないとか、ありえないと聞きます。そこらへんどう思われますか？

次田 素朴な印象として、罵詈雑言が多すぎるのが気になっています……。仲悪すぎでしょう！ これは理論言語学に好意的な側とそうでない側、どっちもどっちだと思います。たとえば、ホーンステインの「コネクショニストの講演を聞いたことがある不運な人であればわかるように」<sup>2</sup>(連載第4回)みたいなのがあったかと思えば、テレンス・セイノフスキーの『ディープラーニング革命』<sup>2</sup>にはフォーダーとかに対する悪口がいっぱい書いてある。こんなふうに長年にわたってお互い罵倒

しあってきたせいで変なしこりが絶対残ってるじゃないですか。研究内容とは直接関係しませんが、野次馬としてはそのあたりの人間臭さが気になってしまいました。

**折田** 我々みたいにバチバチに議論する人たちがいる一方で、「ニューラル的な表現と記号表現のよいところを融合したらいいんや！」と考える研究者もいますよね。自然言語処理の辻井先生が講演でおっしゃっていた記憶があります。これって共闘？ 私はこれめっちゃもやもやするんですが。

**窪田** うん、もやもやしますよね。それで思い出したんだけど、辻井研全盛期の、いかにも辻井研らしい研究の一つって、宮尾さんが博論でやった、HPSG(主辞駆動句構造文法)のちゃんと動くパーザ(parser, 構文解析プログラム)を作った研究なんですけど——これが一番すごい成果だった、みたいなのはもちろん僕の主観にもとづく見解にすぎないですが——HPSGって、ものすごくデータ構造が複雑になるから、パーズングがすごく難しいんですよ。で、あの当時(2000年代)の計算機の性能で実際にパーズングができるようなやり方を、理論的知見と工学的手法の絶妙な組み合わせで編み出して、それを実際に動くものにまで作り込んだ。世界中で誰も成功してなかった時期に。なんて言うのかな、辻井先生は「これとこれをこういうふうによく組み合わせると……」みたいな感じでふわっとしたことを言うのがうまくて。で、実際に手が動いて頭の切れる学生がいると、それが形になるところまでいく。こういう、面白くて難しい問題を実際にある程度解ける形に落とし込む作業に成功するためには、全然別の分野の学問の核心にある本質的な部分をまずそれぞれ正確に捉える必要があって、だけど、それだけじゃ足りなくて、単純なひらめきというか、思いつきを信じてやり抜いてしまう度胸が絶対に必要だと思うんです。そういう感じの共闘ができたらすごく面白いと思います。ただ、そういう困難なことを成し遂げたととしても、そこから直ちに我々側の科学的問いに関して明確な貢献が得られるわけではないのが難しいところで、もやもやするところですよ。

**次田** 岡野原さんも辻井先生からの影響を語ってましたね(番外編1)。その際に、言語学者のAI研究への批判とAI研究者の言語学への批判には相手方に対する理解不足によるものも多いのでは、と指摘されていたのが印象的でした。両方の専門家になるのは難しいとはいえ、情報交換は大事なと感じます。個人的な話ですが、今回AIの本を書いた後に、以前一緒に研究したことのあるNLPerの菅原朔さんから、ステイブン・ピンカーとアラン・プリンスの動詞の過去形に関するニューラルネット批判を解説した箇所を学生に読ませたいっていうコメントをいただきました。ニューラル言語モデルはうまくいってると言われるけど、じゃあ、言語学者たちが発見してきたパターンをどのぐらいうまく捉えられているのか、そもそも言語学者はどのようなパターンを見出してきたのかってことも含めて我々はもっと知るべきだって。

**折田** それは菅原さんだからっていうのもあるんですよ。

**次田** そうかもしれません。そもそもなんで自分が菅原さんと一緒に仕事したことがあるのかっていうと、彼はもともと学部は東大の哲学科で、修士から自然言語処理の方に行ったんですよ。なので、一応彼は僕の後輩になるんです。

**折田** ホーンSTEINはもともと哲学の学生だったんです。何かの問題を考えるときには、抽象的なレベルから具体的なレベルに下っていくのが一番いい、と言っていた。哲学、言語学、その次に計算言語学、自然言語処理みたいな感じで。菅原さんなんて、その良い例ですよ。

**次田** 菅原さんは抽象的な方から具体的な方まで全部カバーしてる、稀有なNLPerなんじゃないですかね。

**折田** 稀有ですね。小町守さんももともと科学史がご専門のはず。

**窪田** 小町さんは東大の教養学部出身で、僕の同学年です。彼も生業として自然言語処理やってるけど、学問的関心はおそらく別のところにあるというタイプの人ですね。

## 言語獲得に生得的知識は必要か？

折田 コネクショニズムの背景にある連合主義について、最近の深層学習の流行をふまえて、どう思います？

次田 連合主義は経験主義の一種でしょうけど、経験主義をもっぱら連合主義として扱うことには疑問を感じています。たしかに、ハップのルールにばかり注目すると、「結局コネクショニストの学習は連合なんですよ」って言いたくなるのかもしれない。でも、たとえば——生物学的な根拠はないとされていますが——逆伝播法のような学習アルゴリズムに「連合」って感じはしないです。なぜそんなに「連合」って言いたがるのかがよくわからないなあ、と。

折田 私が翻訳したホーンステインのブログ記事(連載第4回)やランディ・ガリストルの論文では「連合」「経験主義」「シナプス可塑性」などがひとまとまりの同じグループとして扱われて議論されているように思います。

窪田 それ単純化しすぎな気がするんだけどな。だって、現代的な意味でのDNNって、別に連合主義と何も関係ないような気がするんですけど。アテンションの機構とかあるわけで、そういうのを単に刺激と反応の直接の対応しかない連合主義と一緒にするのはあまりにもアンフェアでは、と思うけど。

折田 ベースになってるのはネットワークじゃないですか。

窪田 それはそうだけど。

次田 窪田さんは鼎談前の打ち合わせでも、「表象・計算・領域特化は自ずと三つ組となり、これらのどれか一つでも否定すると三つ組のすべてを否定することになる」というホーンステインの主張を疑問視されていましたね。

折田 ガリストルの書いた論文を見ると、やっぱりフォードーにすごい影響受けてて。ガリストルはもともと、動物の学習や記憶の専門家なんです。よく「虫でも距離や方向を計算して移動できる。あんな膨大なニューラルネットワークがアリの脳

に入ってるか！」みたいな感じで議論する人なんです。彼が言ってるのは、「シナプス可塑性だけ」では——存在は認めてるんですよ——人間や動物の学習や記憶のしくみは説明できない、ということ。彼は、事象と事象の間のつながりや共起関係だけでは学習や記憶といった認知を説明できない、認知には記号が必要だと主張しています。でも、ネットワークじゃアカン、シナプス可塑性だけじゃアカン、って言うんだったら、代案あるのかって思いませんか。ガリストルは批判するだけじゃなくて、ちゃんと仮説を提案しているんです。それが「細胞固有記憶仮説」というもので、連載第4回で紹介しました。かなりざっくり要約すると、認知のほとんどが、記憶から記号を取り出す操作、これらの記号に対する演算、そしてそれに続く記憶の貯蔵によって構成されるなら、脳は標準的なコンピュータのような構造をもっているに違いない、という仮説です。

窪田 これは僕も面白い仮説だと思って連載記事を読みましたが、言語学の専門家としては、生得的仮説との関連が気になるところです。たとえば、生得的言語知識を想定しない、用法基盤モデルを支持するような研究者たちはガリストルらの仮説をどう考えているのでしょうか。たとえばアデル・ゴールドバーグの構文文法の理論とか見ても、言語が記号体系であること自体、つまり表象の概念は認めているように見えますが、そういう立場と細胞固有記憶仮説は整合的なのか？ 一方で、生成文法は、何らかの意味で抽象的な計算機構っていうのはあるんだということを言い続けている。ただ、それに関しても、たとえば再帰構造とか生成文法的な意味での言語記号に関する「生得的知識」は細胞が「記憶」を「保持」する、みたいな立場で説明できるのか？ いろいろな疑問が湧いてきます。生成文法陣営が主張している「生得的知識」については、昔の鳥の制約みたいな「いかにも」という感じのものに比べて、階層性とかあるじゃないかという程度の話だと、別にそれが自然言語に特有かどうかなんてぜんぜん自明ではなくなってきましたよね。そこらへんでひょっとして

接点がありうるかもしれない。これは確か連載第2回の瀧川さん、岡野原さんの指定討論でも出ていた論点だったと思うけど、ある種の、構造化した知識をうまく処理する機構が人間の脳にはあらかじめ備わっているかもしれない。世界のあり方を把握するために都合がいいので、そういう仕組みが進化の過程で発達した、という可能性は大いにありそうじゃないですか。

**折田** 生得的・領域固有な言語知識を認めない理論による言語獲得モデルの一つに、Chunk-based learning(チャンク基盤学習)というのがあります<sup>3</sup>。音から談話まで複数のレベルでチャンクして(かたまりを作って)、遷移確率などのパターン分布から“Give me X”, “[Verb]-ed”みたいに言語をスロットのあるフレームとして学習していくという仮説です。モーテン・クリスチャンセンとニック・チェイターは、こういう学習方法なら言語に特化した事前知識は必要ないと考えていて、言語獲得は車の運転、楽器の演奏、チェスなどと同じ「スキル学習」の一種だと主張しています。

**窪田** この手のアプローチで、Xバー理論が捉えようとしていたレベルの抽象的な構造的知識を獲得しうる、みたいな話になったらほんとに面白いけど、横目でちらちら眺めている外野としては、まだまだこれからなのかな、みたいな気がします。

**折田** この類の言語獲得モデルに詳しくはないのですが、たしかにこれからの印象です。こういうチャンクやスロットのあるフレームとしての言語知識の上に、より抽象的な言語知識をどこまで仮定しているかはまだ議論の最中だと思います<sup>4</sup>。

話題をガリストルに戻すと、ホーンSTEINやガリストルはシナプス可塑性や神経回路と連合学習を結びつける考え方が経験主義的であると批判していますが、一方で、神経科学の深井朋樹先生は連載第4回の指定討論で、ホーンSTEINの叩いているニューラルネットワークモデルってのは古いんじゃないのかとも指摘しています。哲学者の次田さんからみて、このあたりの議論はどう思われますか。

**次田** 正直なところ、「経験主義は連合主義なん

でしょ」という議論の余地がある前提に立って批判している、ってのが意外でした。ガリストルが書いたものは前に何か読んだと思うのですが、そういう印象はなかったの。

経験主義と生得主義の違いについては、領域一般(domain-general)と領域固有(domain-specific)という対比に訴えるのが個人的にはしっくりきます。たとえば、「条件付け」は連合学習の典型例ですが、これはいろんな領域にも応用が利く領域一般的な学習機構で、基本的には「繋げられるんだっとなんでもいい」。それに対して、チョムスキーは、学習理論が今よりずっと未熟だった1960年代に——今でも未熟と言ってもいいかもしれませんが——「条件付けなんて仕組みで言語を獲得できるわけない。だから言語は学習されるものじゃない」ということで言語獲得装置の必要性を説いた。この構図は今も変わっていないはずで、言語に特化していない汎用の学習機構でいけるのであれば経験主義の勝ち、そうじゃなければ生得主義の勝ち、というふうに自分は理解しています。それを踏まえて、たとえばトランスフォーマーに目を向けてみると、もともとは機械翻訳のために開発されたわけですけど、言語とは全然違う領域のタスクであるはずの画像認識などにも適用できる。この点でトランスフォーマーは領域一般的な学習機構と言える。「連合」ではないにせよ、経験主義には都合がいい。

**窪田** 僕は最近の深層学習の研究の面白い点はそこにあると思っています。チョムスキーが最初に生得主義的な立場をすごく強く打ち出したときは、そもそも proof of concept としてすら、領域一般的な学習装置で言語の処理が可能か、ということがわかっていなかった。そういう時代においてはチョムスキーの議論はやはり説得力が強かったと思います。ところが、そこでしばしば暗黙裡に仮定されていた前提が随分と危ういものであったということ、いま我々が目にしている事態ははっきり目に見える形で暴いてしまったのではないかな。そんなふうに僕としては思うわけです。これを言うとまた、折田さんに「また窪田のEが出た」

と言われると思うけど。

**折田** なるほど。でも、これは何回言っても伝わらないみたいな感じがあるんですけど、ディープラーニングのインプットと人間のインプット全然違うやん！ って。冒頭で少し触れましたが、最近では赤ちゃんにヘッドカメラつけて、その視点から画像と音声を取得して、それをインプットにしてニューラルで学習するみたいな研究もあります。でも、そもそも赤ちゃんにカメラ付けている時点で赤ちゃんの社会認知能力という高い下駄を履いてるんですよね。画像インプットには加工や水増ししてるし、言語データには前処理で解析器かけて分割してたり。こういう点で私は自分自身をすごく“E”だな！ って思うんですよ。「そもそもインプットがぜんぜん違う！」という批判はかなり経験主義的な反論だと思うんですけど。生成文法家が見ている言語知識とか言語獲得の問題っていうのは、やっぱり抽象度や問いの形というのが根本的に違うと思うんです。刺激の貧困の問題は、ディープラーニングとか LLM で全然解決してないですよ。

**窪田** LLM をそのまま、人間の脳に備わっている学習装置の近似として使うのはいくらなんでも乱暴だと。そこは僕も同意します。だけど、その代案はなんなのかっていうのが、わからない。僕の偏見かもしれないんですが、生成文法の理論研究をやってる人たちって、未だにディープラーニングみたいなのがなかった時代の前提にもとづいて、刺激の貧困論法をやってるように見える。そこに強烈に違和感をもつんです。ディープラーニングほど訳のわからないブラックボックスでなくてもいいけど、そこそこうまく設計された学習機構があれば、そこまで事前知識が大きくなくても、限られたインプットから結構習得ができるんじゃないかっていう気がするんです。

**折田** さっき次田さんがおっしゃったような、小規模な言語モデルみたいなものを使ってやってみたら、言語固有ではない領域一般な学習の方法で、ある程度、言語の知識って学習できるかもっていうことですよ。

**窪田** そういう感じですか。その意味で、岡野原さんが本で言っている「言語は学習しやすい特別な性質をもったデータであり、ヒトがその性質を活かした学習を行なえる汎用の学習器をもっているから」という仮説は、わりといい線行ってる直感なんじゃないかなと思うんですよ。

**折田** 一方で、生成文法家が見てるような言語知識は統計的学習のみで本当に学習できる種類のものなのかという疑問はまだ解消されていない。

**窪田** 岡野原さんの受け売りになりますが、言語ってのはなんかわかんないけど習得しやすい変な形をしているんだと思うんです。で、生成文法の人たちは今まで何をやってきたかという、その、変な形というのが、具体的にどういう形かというのを職人技的な直感で鋭く捉える作業を過去 70 年ほど延々としてきたのだと思う。まだ我々が解けてない問題っていうのは、子どもがその変な記号体系に関するごく限られた情報を入手するだけで、一見すると複雑に見えるシステム全体をサッと習得できてしまう、そここのところの具体的なメカニズムですよ。僕は、生成文法の一般的な立場に比べると、そこで学習機構の果たす役割がもっと大きいんじゃないかと思ってるということです。折田さんの方に僕が「歩み寄り」とすると、生成文法の人たちが「事前知識」という言葉で言っているものは、僕の捉え方だと「学習機構の設計」と言った方がよい、ということになるのかも。

**折田** それはそうかもしれない。生得的な言語知識と言ってるものは、実はすごく範囲が狭くて領域一般な学習でできる可能性もある。

**窪田** そうそう。学習機構でしかも領域一般な部分が実は結構大きいんじゃないかという疑惑。

**折田** そこは実際に言語獲得の研究者がまさに研究している問題ですね。領域固有しかないと思ってる人はほとんどいないと思う。

**次田** 人間の子どもが短期間に言語を獲得するのは、言語そのものが学習しやすいという性質をもったデータだからだ、という岡野原さんの仮説ですが、似たような発想は前からあると思います。たとえば、テレンス・ディーコンの『ヒトはいか

にして人となったか』<sup>5</sup>。言語が大昔のヒト属によるある種の発明品だとすると、それは頑張っただけで覚えなきゃいけない。その際、他よりも早く学習できる者が生き残ってより多くの子孫を残したことで、次第に短期間で習得できるようになっていったという可能性が一方にはある。ピンカーのような生得主義者はこの路線ですね。だけど、言語が発明品であるなら、「覚えやすい形で」使われるように発明・改良されていった、という可能性もある。人間の脳に適した言葉の並びしか使わないようにするみたいな。ディーコンはこっちの路線なんですけど、二つのシナリオは両立するように思えます。効率よく言語を獲得できるように脳の配線が変わっていく一方で、アウトプットとして出てくる言葉の並びに関しても、人間がより使いやすいものが特に好んで選ばれる、といった好循環があったのかもしれない。これは現実味のあるシナリオだと思います。余談ですが、「学習しやすさ」ってなんだかアフォーダンスみたいですね。言語そのものが「学んでくれ〜」っていうメッセージを放っているのかもしれない。

## LLM と概念の表象

次田 LLM は人間の脳とはまったくの別物ではないか、という論点とは別に、LLM の限界を示唆する現象もいろいろと指摘されています。代表的なものとしては、まず、ハルシネーション。学習データに含まれていないことや、学習データと矛盾することを言うてしまう。この問題は Chat GPT などを使ったことがある人はみんな経験している。もう少しマイナーなところでは、質問応答や含意関係認識などのタスクにおいて、標準的なベンチマークではうまくいったのに、同義語で置き換えたものを問題すると正答率が悪化する、といった欠陥。それから、問題文を読ませて4択で答えさせるタイプの質問応答のタスクで、選択肢しか与えられてないのになぜか正解しちゃう、みたいな研究もありますね。無茶苦茶でしょう、それって(笑)。

折田 & 窪田 (爆笑)

次田 いったいどういうパターンを学習したんだか……。こんな感じで、LLM のおかしい挙動はいろいろある。じゃあ、そのおかしさは一体どこから来ているのか。「AI には身体がないからダメなんだ」とか言う人をときどき見かけますけど、自分はこの種の批判がさっぱりわからないです。カメラやモーターを LLM に備えつけたからといっていま挙げたような問題点が根本的に改善するとは思えないので。

窪田 それはそうですね。でも、選択肢しか見なくても答えられるっていうのは、たとえば、受験テクニックを極めた学生とかだったらやってるかもしれないことですよ。

次田 はい、それはその通りだと思います。LLM のおかしい挙動を笑うのはある意味簡単なんだけれども、その批判の矛先は人間にも向かいうる、ってことですよ。LLM が嘘つきだとしても、人間にだって常習的な嘘つきや陰謀論者はいる。LLM は問題文を読まずに選択肢だけで正答するかもしれないけど、同じようなテクニックは受験生も使っている。LLM は同義語で置き換えただけでパフォーマンスが悪化するかもしれないけど、人間にだって同義語によって言い換えられると混乱してしまったり、逆に、単に言い換えただけなのに、それまで理解できなかったことを理解した気になったりもする。こんなふうを考えていくと、言葉を理解するとか、概念を把握するとはどういうことなのか、わけがわからなくなってくる……。

折田 次田さんが言った「AI には身体がないからダメなんだ」という批判、私もよくわからない。身体性に関していえば記号接地問題もわからない。AI やロボットの記号接地と人間の記号接地が平行で議論されているのを見かけますが、「カップ」が脳で概念としてどのように在るのか、どのように記憶として維持されているのか、こんな問題がほとんどわかっていない中で記号接地をどう考えればいいのかという素朴な疑問が。

窪田 むかし認知言語学とかで流行ったプロトタイプ理論みたいな、ああいうのではダメなんですか？ 典型例があって、そのコアなものだったら

備えている特徴の束があり、典型例から外れているものは全部の特徴を満たしていない、みたいなやつ。たとえば「鳥」って言ったなら「カモメ」とか「スズメ」は典型例で、「ペンギン」はおそらくプロトタイプを中心からは外れるけど「鳥」。この手の概念観は記号主義的な立場よりはニューラルネット的なものに親和性が高いですよ。まあ、僕も概念の獲得みたいなものの研究はほとんど知らないですけど。

次田 プロトタイプ理論に関していうと、たぶんフォーダーは最大の批判者の一人ですよ。彼のポイントはただ一つ。合成原理を満たさないから、ということです。たとえば、「pet」の典型例は犬や猫で、「fish」の典型例はマグロとかマスとか。でも「pet fish」の典型例はグッピーだ、と。「pet」と聞いて我々がまず犬や猫を連想するというのはおそらく事実ですし、こうした連想のパターンを捉える心理学の理論は必要ではあるのでしょう。でも、それがPETやFISHやPET FISHといった概念の中核にあるのかと言うと、自分はそんな気はしない。だって、「pet」の典型例と「fish」の典型例からどう計算したら「pet fish」の典型例を導き出せるのか全然わからないです。フォーダーなら、PET FISHとはPETかつFISHであるような概念だ、という具合に簡単に合成できると言います。「ほら、合成性って大事でしょう？　だって我々は無限に多くの言語表現を作って、それぞれの意味がちゃんとわかるのだから。そのためには、単純なパーツから全体の意味を組み立てられるようになってないはずいでしょ？」という話になる。もちろん自然言語の語彙にはイディオムもありますけど、それでも多くの表現に関して、「パーツから全体の意味を計算できるべき」という発想は生き残ると思います。

一方で、フォーダーにもやばいところはある。概念獲得に関するPrimitivismと呼ばれる立場で、要するに概念は「最初からある」と。もちろん、カップが存在しない世界に住んでいる人々はCUPという概念をもってないだろうけど、素質はもっているんで、環境からの刺激をちょっとで

も受けたら、何かのスイッチが入るといふ(笑)。「刺激の貧困」の極致みたいになっている。Primitivistはここで免疫系の比喩に訴えるんですよ。我々は生まれた時点ですでに、あらゆるタイプのウイルスに対応できるくらい多様な抗体をもっているんだけど、そのうちのいくつかだけが残る。それと同じように、我々は生まれたときから大量の概念をもっているんだけど、そのうちのいくつかだけが刺激を受けて発現する、と。

折田 え、それマジで言ってるんですか？

次田 マッシモ・ピアテッリ＝パルマリーニとかはマジで言ってました。フォーダーもたぶんそうです。それはないだろうって思いますけど、でも今回AIの本を書いてみて、「自分は思っていたよりフォーダーが好きだったんだな」って感じました。好きな部分はやはり合成原理ですね。プロトタイプ理論は面白いと思うけど、これが概念だと言われるとなんか違う。というか、ステレオタイプとかプロトタイプって合成原理と相性が悪いんですよ。たとえば「外科医」って聞くと「男性」を連想するという話があるじゃないですか？典型的にはそうだから。でも、いくら我々が心理的にはそう連想しがちだといっても、「外科医」という概念の中に「男性」は含まれないでしょう。そうじゃないと「女性の外科医」って表現が意味をなさないので(笑)。

窪田 それは面白いですね、うん。

次田 LLMがあやういのは、ステレオタイプとかプロトタイプとかそういう情報を全部込みで学んでしまうところだと思います。形式意味論を勉強したことがある人なら、「合成性は大事だ」という直感を叩き込まれていると思うんです。そういう人からすると、「外科医」にさまざまなステレオタイプがまわりついているにしても、それはこの概念の中核にはない、と言うことに抵抗はない。むしろそう言いたいはずだと思うんです。でもLLMはそういう腑分けを全然しないで、テキストコーパスにあるパターンを全部取り込んでしまう。表面的には人間の言葉遣いに近いものが出てくるけれども、やはり言語学者は不満をもっ



て然るべきではないかと。

窪田 はい。これがまさしく我々言語学者が守らないといけない「最後の砦」だと思います。言語ってやはり変なバイアスがいろんなレベルで染み付いてて、そういうものを記号体系として我々は使っているっていう事実は動かしがたい。で、少なくとも建前としては、形式意味論や統語論はそのバイアスの染み付き方が問題にならない部分や、バイアスの染み付き方に一定の制約が観察される部分の規則性を扱っている、ということです。一応理論言語学の人間として、ここははっきりさせておきたい。ただ、結局そういうのを全部切り捨てたから形式意味論とか統語論とかってなんか象牙の塔の学問みたいなものになってしまったように思うんです。で、分布意味論やLLMは形式言語学が切り捨てた部分をうまくブラックボックスにして、実際に実用的なタスクでうまく動作する機械として実現してしまった。そこに個人的にはジレンマを感じます。爆弾発言かもしれませんが、「研究者」という鎧を脱ぎ捨てたところの僕の本音としては、形式言語学側の建前ってほんとに完全に擁護できるのだろうか、みたいな気分になることもあります。

## ふたたび、「共闘」の可能性について

折田 窪田さんはなんだかんだ言ってLLMを研究に使いたいんですね。

窪田 使いたいですよ。でもうまい使い方をまだ探し当てていないというのが正直なところ。

折田 なんでニューラルネットワークじゃないとだめなんですか？

窪田 LLMやニューラルでなくてもいいけど。なんでニューラルを使うかという、他と比べたら一番性能が出そうに見えるのが主な理由で、あとは機械学習のアルゴリズム自体には僕は興味がないので、単に皆が使っているからという身も蓋もない理由。今までは実テキストのパージングで、理論言語学者である僕なんかに興味をもつような、ある程度複雑な意味解析のタスクをやろうとすると、全部rule-basedで書こうとして失敗すること

が目に見えているというような時代がずっと続いてきたわけです。ところが、ニューラルの時代になって、そのハードルを克服する便利な道具が意外と簡単に手に入るようになった。やったぜ、これはチャンスだ！みたいな気分には僕は今なっていて。たとえば、内容語の意味を部分的に参照するような構文・意味解析のシステムを作る研究を今少し進めているところなんですが、そういう場合、そこら辺に転がってるトランスフォーマーのモデルをちょっとファインチューニングして使う、みたいなことが割と簡単にできて、しかも結構精度がいいわけです。その際に「外科医」って言われたら「男性」を連想するみたいな、変な挙動をするモデルだったとしても、僕が解こうとしている問題に対してはその手の問題は無視してもいいノイズであることが多い。そういう感じで割り切って使う、ということですね。決してそのモデル自体が人間の認知のモデルとしてそのままの形で役に立つとは思っていません。

折田 私はニューラルネットワークを使った言語獲得の研究は全く買えないと思っています。説明力がないし、そもそも人間の認知モデルとして全然妥当じゃないと思っています。そんなモデルから出てきた結果や知見は一体何なんでしょう？

安西祐一郎先生は「認知アーキテクチャの一部に、説明力の不足している生成AIを組み込むと、得られるモデルはむしろ認知アーキテクチャの説明力を減じることになりかねない<sup>6)</sup>と最近の論文で書かれてました。

窪田 折田さんが冒頭で出していたワイ・キン・ヴォンらの研究なんかまさしく典型例ですね。

折田 次田さん、このあたりのことについてどう思いますか？

次田 コネクションニズムについてさんざん批判的なことを言っておいて、最後にちゃぶ台をひっくり返すようですが、記号主義とコネクションニズムのどちらも部分的には正しいんだろう、とは一応思っているんです。本当の問題は、両陣営の洞察や強みを合体させることの難しさにあるのでしょうか。単なる哲学者の自分にこの方面で貢献できる

ことがあるのかはわかりませんが、野次馬として最近の研究動向を追っている限りでは、たとえば、数学オリンピックの幾何学の問題を解く AlphaGeometry などは面白いと思いました。補助線を引く、といった数学者の直観に相当する部分をトランスフォーマーが担当し、厳密な証明は古典的な推論エンジンが担当する、という合わせ技で人間の優勝者レベルの成績を取めた、という話です。

それはそうと、今回話をうかがっていて、強すぎる経験主義がデータ万能主義の風潮を招いている、という折田さんの懸念には共感する一方で、実装レベルの忠実性には自分はそこまでこだわりのないのかも……とも感じました。たしかに、LLM が人間の認知モデルになるとは思いませんし、AlphaGeometry だって人間の数学者とどこまで似ているかは怪しい。でも LLM が便利な道具であることまでは否定できない。最終的に知りたいのは人間であって、人間のデータこそが証拠なのだとしても、LLM は簡易版人間くらいにはなるかもしれない。だから、人間の被験者を集めるのはお金もかかるので、いきなり人間試す前に、とりあえず LLM に聞いてみて、「こういう現象がひょっとしたら本当にあるのかも」となったら人間で調べてみるとか。それがうまくいくかはわかりませんが、あくまで発見的な使い方をするということなら、そこまで違和感はないです。

**折田** なるほど。そういった意味では、粗いかもしれないけど近似にはなっているし、研究によっては使ってみてもいいのかもしれないね。まあ人間の実験もそんなに綺麗なデータではないですからね。でも言語学習のモデルとして使うのなら、また A4 一枚で批判ぎっしり書きちゃうかも(笑)。  
**窪田** そろそろまとめましょうか。といっても、まだまだ全然話し足りないですね。AlphaGeometry と赤ちゃんの頭部にカメラ付けてデータ取っの類の研究はどこが似ていてどこが違うのか、こういうのは自然科学・人間科学の研究としてありなのか、みたいなことはもっとちゃんと整理してみたい。「共闘」は可能か？ の話に結局戻り

ますが、たとえば、その折田さんの A4 一枚ぎっしりの真っ向からの批判の一つずつに、モデリング陣営が真剣に応答するようになれば、それが「共闘」なんじゃないでしょうか。単に口汚く罵り合うのでなく、真面目に丁寧に真剣勝負する。最近亡くなったダニエル・カーネマンが adversarial collaboration(敵対的協力)という言葉でそういう共闘のあり方を提唱していたらしいです。言語と認知の研究の世界でこの意味での共闘が可能であれば、そこに参戦して闘いつづけていきたいな、と決意を新たにしました。

(2023年12月27日、早稲田大学にて収録後、加筆修正してまとめ)

#### 文献

- 1—辻井潤一: 人工知能学会誌, **27**(3), 273(2012)
- 2—テレンス・J・セイノフスキー: 『ディープラーニング革命』, 銅谷賢治監訳/藤崎百合訳, ニュートンプレス(2019)
- 3—M. H. Christiansen & N. Chater: Behavioral and Brain Sciences, **39**, e62(2016)
- 4—K. Abbot-Smith & M. Tomasello: The Linguistic Review, **23**(3), 275(2006); B. Ambridge: First Language, **40**(5-6), 509(2020)
- 5—テレンス・W・ディーコン: 『ヒトはいかにして人となったか: 言語と脳の共進化』, 金子隆芳訳, 新曜社(1999)
- 6—安西祐一郎: 人工知能, **39**(2), 187(2024)

タイトル画像クレジット: vladystock/123RF

\* 次回(11月号掲載予定)は言語獲得と文法変異がテーマです。本連載の最終回となります。

**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)

**窪田悠介** くぼた ゆうすけ  
国立国語研究所准教授(統語論・意味論)

**次田 瞬** つぎた しゅん  
富山大学人文学部講師(心の哲学・言語哲学)