［連載］人間の言語能力とは何か――生成文法からの問い❷

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

ノバート・ホーンスティン
折田奈甫・藤井友比呂・小野 創〈編訳〉

人工知能(AI)という分野が謙虚であったことなど一度もない。はるか昔の私が大学院生だった頃から，なんどかの「AI ハイプ・サイクル」(訳者注：新しい技術に対する人々の期待値の変動)を個人的に経験しているが，その称するところ，どの回も前回の技術より画期的で，前回の失敗に対する不満や失望に覆われた「冬の時代」を乗り越えて発展したという。私はといえば，ただこれらの盛衰を傍観していたのではなく，初期の人工知能モデルに関する批判的議論を寄稿したこともある[1]。とはいえ，今回のこの盛り上がりはある重要な一点においてこれまでとはかなり異なるかもしれない。最近の ChatGPT のような大規模言語モデルは，何か楽しくてびっくりするようなことができるようだし，我々の日々の生活に技術的な恩恵(あるいは邪悪な何か)をもたらすことさえあるだろう。いずれわかる。そして，これらの技術から便利で実用的な用途を見出せるかについて，私は極めて寛容かつ柔軟に考えている。

私が極めて非寛容になるのは，ChatGPT のような大規模言語モデルが，人間の認知モデルとして，特に言語という領域において，役立つ可能性があるかを考えるときだ。これまでの他の全ての「ブレイクスルー」は，その人工知能技術の長所らしきものを誇示するだけで不満足な結果にとどまった。開発者たちは，彼らの作った人工知能モデルが人間の言語能力の理論として革命的なブレイクスルーだと即座に触れ回る衝動を抑えることができなかった。今回も科学者としてこれを真に受けてはならない。ChatGPT の振る舞いを少し見るだけで，またしても正しい認知モデルにはならないだろうとわかる。説明しよう。

70 年以上にわたる綿密な研究によって，言語学者は人間の言語能力の内実についてかなり多くのことを知っている。特に言語学者は，自然言語の文法が，これらが言語能力の産物であるがゆえに，ある特定の形式的な性質をもっていることを発見した。自然言語の文法とは，ある特定の性質をもつ規則を含み，他のありえる規則は含まないのだ。さらに言うと，言語能力は，ある特定の言語のインプットを与えることで，その言語固有の文法規則を生成する。インプットからこれらの文法を学ぶには証拠が乏しいにもかかわらずだ。言語学者が 70 年間にわたって言及してきたように，文法に関する知識の獲得は，学習者である子どもにとって相当に少ない観察データしかないにもかかわらず進んでいく。言語獲得については別のところで議論をしてきたので，この点についてここで長々と述べるつもりはないが，言語獲得の問題は，ChatGPT が(その技術的長所が何であれ)，3 つの論点において人間の言語能力の理論にとって実質的な(つまり多少なりとも正しい)基礎や根拠を与えないということを明らかにするのに役立つ。

第 1 に，ChatGPT は人間の母語話者のようには文の容認性を「判断」できないようだ(実のところ ChatGPT は判断など全くしないので，皮肉を込めてカギ括弧付きの「判断」)。ハワード・ラズニック(Howard Lasnik)の退職を祝うワークショップで，ルイージ・リッツィ(Luigi Rizzi)が自明で単純な例を紹介していた(訳者注：ラズニック，リッツィともに統語論が専門の生成文法家)。リッツィは，文中の代名詞が何を指示できるかという質問に対して，ChatGPT がどのように答えるかを試した。ラズニックがずっと前に指摘していたように，以下の(1)のような文で，代名詞の he と固有名詞の John は同一人物を指示できない。これは，he と John が同一人物を指示できる(2)のような文と対照的だ。

(1) **He** said that **John** is a nice guy.

(彼はジョンがいい奴だと言った。※「彼」がジョンを指示する解釈はできない)

(2) The people that **he** talks to say that **John** is a nice guy.

(彼が話しかけている人々はジョンがいい奴だと言う。※「彼」がジョンを指示する解釈も可能)

リッツィが言うように，この 2 文のような解釈の違いは 4 歳児であっても気づくようなことである。しかし ChatGPT はそうではなさそうだ。ChatGPT は，(1)と(2)のどちらにおいても，he と John が同一人物を指示する解釈が不可能だとみなす。つまり，もし ChatGPT の応答を同じような質問を人間にしたときの反応と同様に扱うのであれば，このモデルは人間の大人や子どものように代名詞の解釈をしていない。

次に，リッツィはこの ChatGPT のふるまいの原因を分析する。ChatGPT の応答は，人間であれば決して身につけないような種類の文法規則に従っているようだと述べている。その文法規則は，he のような代名詞と，その代名詞より後に来る John のような名詞表現が，同一の対象を指示するのを禁止するというものだ。それに比べて人間は，同一対象を指示する可能性のある先行名詞表現よりも代名詞が階層構造上より高い位置にあるとき(訳者注：厳密には，生成文法の用語で「代名詞が名詞を c 統御する」と言う)，同一対象を指示する解釈を受け付けない。つまり，ChatGPT の応答を文字通りに受け取るとすれば，人間と ChatGPT の振る舞いは単純な例であっても異なるのだ。二者の反応の裏にある文法規則は質の異なるものであり，ChatGPT が使っている規則は人間の言語ではありえない規則だ。言い換えると，その間抜けな応答を見るに，ChatGPT は，人間の場合は言語能力によって学習が不可能になっている，ありえない文法規則に従っているようだ。

さらにこうも言える。チョムスキーが最近の記事[2]で述べたように，人間にとっては不可能な文法規則を ChatGPT が「獲得」できるがゆえに，これらのモデルは人間の言語能力について大した説明にはならないのだ。よって，たとえ ChatGPT が人間の言語のように階層構造に依存した文法規則を獲得できたとしても(上記の例ではできていないが)，階層構造に依存しない規則も獲得するのであれば，言語能力がまさにこのような規則の学習を禁じるという点において，人間の言語能力の理論としては依然としてお粗末なのである。

3 つ目の論点を考えてみよう。人間の子どもは，母語を獲得する過程で数百万程度の文を聞く。膨大な量のデータに思えるが，子どもが最終的に到達する文法知識を考えるとそうでもない。まず，子どもが耳にする言語データは曖昧であり，最終的に獲得する多くの文法知識に関係するデータを含んでいないという点において不完全である。データの偏りもある。たとえば，“The man who I saw said that Mary likes to eat bananas that are ripe(私が会った男はメアリーが熟したバナナを食べるのが好きだと言った)”のように複数の埋め込み構造をもつ文を子どもがたくさん聞く可能性は低い。“The pig is tickling the hen(ブタがめんどりをくすぐっている)”のように空想上の出来事を表す文がたまに出てくるとはいえ，子どもが聞く文で典型的なのは短くて埋め込みがなく形の整った文だ。さらに言うと，これらの典型的な文は他の意味で偏りがある。たとえば，ある言語の動詞や名詞の活用変化形の全容がわかるような用例が揃ってデータに出てくるわけではない(子どもの言語インプットがいかに偏りがあり不完全かを知るには Yang[3] の素晴らしい議論を参照のこと)。

何はともあれ，言語学者と心理言語学者は，正しい帰納推論に子どもを導くためのデータが不十分なのにもかかわらず，文法の獲得が起こることをずっと前から知っている。つまり，子どもの言語獲得は，ChatGPT などに代表される大規模言語モデルによる言語データの学習とはずばり真逆の条件下で行わ

れているようなのだ。ChatGPTでは，3000億もの単語がインプットとして使われていて，さらに，子どもが普段接するような文とは比較にならないくらい複雑な構造をもつ文がたくさん含まれている。電子化された書籍，様々なWebページ，ウィキペディアの記事など，ウェブ上に存在するテキストがデータとして使われている(すなわち，埋め込みのない単純な文ばかりではないのだ！)。つまり，子どものインプットの場合は数百万程度でほとんどが単純な構造をもつ文が占めているのに対して，ChatGPTの場合は，1000億単位の単語数であらゆる種類の語や文構造を含んでいるのだ。要するに，ChatGPTのようなモデルの学習がどのようなものであれ，子どもの言語獲得の問題とは大雑把な記述レベルでも全く別物なのである。

そして最後に見ていきたいのは，あらゆる有限集合(それがどれほど巨大であっても)は無数の方法で一般化が可能であり，にもかかわらず，どの一般化でも最初に観察したデータをとらえることができるという問題だ。この帰納の問題は，膨大な量のデータがあったとしても払拭できない。つまるところ一番大きな問題というのは，人間が有限の言語データの集合から，ほかでもない特定の方法で一般化を行うということだ。なぜか？　これこそが難問であり，より多くの学習データがあればうまく解決できるわけではないのだ。この「一般化問題」は，言語能力を用いた推論過程において，ある特定のバイアスを導入することによってのみ解決できる。つまり，言語学者が言語能力の詳細を特定するときに行うような推論をすればよいのだ。我々が知りたいのは，ChatGPTのようなモデルがどのようなバイアスを組み込んでいるのかであり，これらのモデルの能力がどのように人間のそれと似ているかを考える上で，これこそが知るべきことなのである。

さて，本稿をどう結論すべきだろうか。ChatGPTのようなモデルが人間の言語に関する認知モデルになりうると本気で考えている者は，数千億単位の言語データではなく，子どもが言語獲得の過程で経験するような数百万程度の単純で不完全なインプットを用いて(つまりウェブから取れる限りのテキストデータを使うのではなく)これらのモデルを調査する必要があるだろう。加えて，これらのシステムに組み込まれているバイアスがどのようなものなのかも特定する必要があるだろう。今までのところ，これらは成し遂げられていないし，期待はできない。ほんの少しのデータで訓練されたシステムが，技術的に価値あるレベルまで達する可能性は低く，うまくいかずに大失敗となるのは目に見えている。さらに言うと，熱心な支持者たちから総じて聞くのは，今のところChatGPTが内部で何を行なっているのか，どういうバイアスをもっているのかは不透明ということだ。よって，ChatGPTのような大規模言語モデルが人間の言語能力のモデルとして科学的に妥当かを考えるとき，現在のこれらのモデルは筋違いだと言わざるを得ない。

ChatGPTは娯楽として何時間でも楽しめる(私も楽しんでいる)。どういう応答が返ってくるのかを見るのは実に楽しい。過去のAIハイプとは違い，今回は技術的に実質的で重要だと評価されるだろう。とは言え，上述した3つの論点において，これらの技術が人間の言語能力に関する科学的理論として価値あるものになるかは大いに疑問だ。もちろん，私が間違っているかもしれない。でも間違っているとはどうしても思えない。いずれにせよ，これらのモデルが人間の言語能力の科学的説明になり得るかを考えるとき，今回のAI革命が過去のよりもうまくいくと考える理由は今のところない。

（翻訳：折田奈甫）

文献

1—B. E. Dresher & N. Hornstein: Cognition, **4**(4), 321(1976)
2—N. Chomsky: "The False Promise of ChatGPT." The New York Times, March 8(2023)
https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html
3—C. Yang: Significance, **10**(6), 29(2013)

【指定討論 1】

言語モデルは単純な学習則で複雑な推論を実現する

岡野原大輔

ChatGPT やそれに類した技術(まとめて CGPT とよぼう)は，次の単語を予測するという単純な学習則のみで，人と同じように話し，言語の背後にある概念をあたかも理解しているようにみえる。

次の単語を予測するという目標を達成するため，副次的に単語や文の意味やそれらの合成則を獲得する，いわゆる自己教師あり学習によって文を読めたり書けたりできるようになることがわかっている。また，次の単語を順に生成するだけでも複雑な推論も実現し，任意のチューリング計算機を模倣できることがわかっている[1]。

このように CGPT は高い能力を備えているが，私は，CGPT が現段階では人の言語能力そのものを説明するものとは考えていない。この点ではホーンスティン氏の主張と一致している。しかし，CGPT がなぜこのような単純な学習則で，複雑な言語処理を実現できるかを理解することは，人の言語能力がどのように実現されているのかを理解するために部分的にであっても役立つものだと考えている。

そして，言語は脳の中で計算され，情報処理をしているという面で，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。

筆者は現在，機械学習や深層学習の研究開発や実用化を進めている，いわゆる「経験主義側」の人になるだろう。一方，20 年前の大学時代に所属した研究室では，生成文法系の言語学にもとづいた文法理論などを用いた自然言語処理を実現しようと取り組んでいた研究者が周囲におり，いわゆる「生得主義側」の考え方にも多少なりとも触れたことがある。

CGPT は Transformer とよばれる計算モデルを使って，それまでの文章に後続する次の単語を予測できるように学習するのみで様々な能力を実現した。学習の際は，様々な言語現象を捉えるための仕組みを明示的に利用していない。

例えば，従来の機械翻訳の実現には高度な構文解析や意味解析を組み合わせ，複雑な推論規則を学習させる必要があった。それが，CGPT は特別に翻訳用に学習しなくても次の単語を予測するだけで実現できる(翻訳文が学習データ中に一部含まれているがほんのわずかである)。同様に，要約をしたり，抽象的，仮想的に考えたり，他のツールを扱えたりする能力が次の単語予測だけで獲得され[2]，一定以上のスケールの学習で自然に発現する。

CGPT はブラックボックスではなく，なぜそのような能力を獲得できているのか，できていないのかは急速に解明されつつある。現在わかっていることを以下に示そう。

言語やプログラムのように構成性をもち，再帰を扱わないと解けないような問題に対し，CGPT はそれを Transformer のモデルを使って勾配降下法を使った学習によって，構成則や再帰則を獲得して解くことができることが理論的に証明されている[3]。ただし，扱える再帰数や問題の複雑さには限界があり，特に小さいサイズのパターンは丸暗記した上でそれを適用している場合がある。

ある文脈自由文法から生成された単語列に対し，次の単語予測学習を行うだけで，その文法を同定し獲得することができ，さらに，学習に用いる単語列がその文法の規則から多少外れていても，文法を獲得できるという頑健性を備えている。また，内部状態は係り受けや，CFG の内部ノードと対応していることもわかっている[4]。

なお，CGPT が登場する以前にも，ニューラルネットワークの学習の過程で単語の概念がどのように獲得されるかについて多くが解明されている。例えば Saxe らは[5]，複数層のニューラルネットワークを使って単語の共起関係を予測できるように，勾配降下法を使って学習した場合に，単語の概念はちょうどよい抽象度の概念からまず獲得され，次にそれらが詳細化あるいは抽象化されていく(例えば犬というちょうどよいレベルの概念を最初に獲得し，次に柴犬や哺乳類といった概念を獲得する)。そこでは観測するデータの順序や内容が変わっても，同じ概念を同じ時期に獲得できることが示せ

る。

また，人の脳の場合は，一時的にシナプス重みを変えることによって短期記憶を実現したり，文脈に応じて挙動を変えたりすることがわかっているが，それと似た機能をCGPTはTransformer中の自己注意機構によって実現している[6,7]。

このように今のCGPTは特別な仕組みを導入しなくても単純な学習則で，同じ単語の概念や文法を獲得できることがわかっている。また，単語の概念や文法は前もって与えなくても，学習によって共通した概念や文法が得られることが計算機上で再現できている。

こうしたことをふまえて，ホーンスティン氏の3つの論点について考えてみよう。

例文(1), (2)の共参照については，CGPTによる反応だけをみても共参照を正しく処理できているのかはわからない。CGPTが内部でどのように処理しているかを解析できる手法は登場してきているので，これらを使ってCGPTが内部で様々な文法処理を人間と同じように処理しているのか調べる必要があるだろう。

その次の，CGPTが不可能な文法規則を獲得しうるから，その学習は人間の言語獲得の仕方と違うという論点については，学習可能性と，実際にそのモデルを学習するかは違うことに注意してほしい。一般にニューラルネットワークや他の多くの機械学習モデルは任意の関数を表すことができ(普遍性定理)，あらゆる規則を獲得しうる。その上で学習の結果，どのような規則を獲得できるのかはどのようなモデル，学習則を使うか(帰納バイアス)，学習データを使うかによって変わる(例えばベイジアン的な解釈は文献8を参照)。CGPTが言語データという学習データを使って学習した場合も，学習可能性で議論するのではなく，実際にどのような文法規則を獲得するのかを議論する必要がある。特に，文献5の例でもあげたように獲得される概念やそれらの組み合わせや処理の仕方は学習の仕方によらず，データ自体の統計的な性質に従ってのみ決まる。言語データ自身がもつ特殊性と，モデルや学習則による帰納バイアスによってCGPTでどのような規則が獲得されるかが注目される。

第3の，大量の学習データ量を必要とするため，人とCGPTが違うという論点は，筆者もその通りと考える。人間の学習の方が圧倒的に効率が良いし汎化性能も高い。この学習効率の良さは言語だけでなく画像認識や音声認識などでも同様である。この点では明らかにCGPT(や今のAI)と人の言語能力は別であるといえる。

私は，人が効率的に学習できるのは，生まれてから成長する過程で様々な学習が決まった時期に自動的に発動することでブートストラップ的に学習する，つまり獲得した学習結果を他の学習シグナルに使うことで効率的に学習しているからではないかと考える。これらは生得的であり，脳の発達過程などで決められたタイミングで決められた領域間で組織が成長したり枝刈りされたりすることによって実現されている。こうした点は今のCGPTにないが，これらの知見がとりこまれてデータ学習効率を上げていくことが進められるだろう。

最後に今後について少し述べる。他分野でもみてきたようにAIは驚異的な速度で進化し，CGPTによる言語生成や言語処理も今後10年でさらに改善される可能性は高い。これらは人の言語処理とは異なるものになると思われるが，これらでわかってきた処理方法から人の言語処理や思考を探る上で参考になる部分もでてくるだろう。人以外の言語を話せる存在によって，人の言語処理も新たな形で理解できるのだと考えられる。

文献

1—E. Malach: “Auto-regressive next-token predictors are universal learners.” arXiv:2309.06979

2—S. Bubeck et al.: “Sparks of artificial general intelligence: Early experiments with GPT-4.” arXiv:2303.12712

3—M. Hahn & N. Goyal: “A Theory of emergent in-context learning as implicit structure induction.” arXiv:2303.07971

4—Z. Allen-Zhu & Y. Li: “Physics of language models: Part 1, context-free grammar.” arXiv:2305.13673

5—A. M. Saxe et al.: PNAS, **116**(23), 11537(2019)

6—J. von Oswald et al.: “Transformers learn in-context by gradient descent.” arXiv:2212.07677

7—R. Zhang et al.: “Trained transformers learn linear models in-context.” arXiv:2306.09927

8—A. G. Wilson & P. Izmailov: “Bayesian deep learning and a probabilistic perspective of generalization,” NeurIPS 2020(2020)

【指定討論 2】

なぜ経験則は説明の論理として受け入れがたいか

瀧川一学

高橋悠治は対談「他者の痛みを感じられるか」において，懸命に答える努力をしていた対談相手に「つまり，質問があれば答えがあるというふうに思われるわけですか」と問うた*1。問いがあることは答えの存在を保証しない。この意味で，私は「言語能力」の理解が言語によって正しく表出できるという前提自体に極めて懐疑的である。その上で，ホーンスティンの記事の根底に私が感じた標題の問いについて懸命に答える努力をしたい。

ホーンスティンはChatGPTなどの生成言語モデル(generative language models)が有用で楽しい技術だと認めつつ，「言語能力」の理解に資するかに関して，3つの論点で懐疑を呈している。私はホーンスティンとは少し異なる理由で同様の結論，つまり「生成言語モデルは人間の言語能力の有用なモデルとは言えない」という主張を支持する。同時に，二項対立にも見える生成文法と生成言語モデルは，ある観点では驚くほど似ている点を指摘する。こうした整理の上で，生成言語モデルが具現化した奇妙な言語の在り方――呪文の体系――とその受容について私見を述べる。

1つ目と2つ目の論点は，代名詞の共参照についてである。生成言語モデルは次の単語を予測する確率分布であり，生成文はそのサンプルの一例に過ぎない。したがって，一般に「○○ができない」と示すことは大変難しい。例えば，「A is B」という文で学習されたモデルが「B is A」と応答できるかを問うReversal Curseの検証[1]では，GPT-4はWho is Tom Cruise's mother?という質問には79%正解したが，逆の質問Who is Mary Lee Pfeiffer's son?には33%しか正解しなかった。論理的一貫性は担保されず，前者ですら常に正解するわけではない。その上，言語モデルの応答は，使用モデルや使用言語，文形に強く依存する。GPT-4での結果は入力文の形や言語によって変動したが，正答する確率の方が高かった。何より，こうした文法的な問題は技術的に容易に対応されうる。例示できる限り，単に新たな事例データとして追加できるからである。また，例で挙げられたc統御のような構造的関係を生成言語モデルが使用していない，と示すことも同様に難しい。生成言語モデルはプログラミング言語の扱いを最も得意とし，文法規則が既知で厳密な人工言語にも事例のみで応答できる。次の単語の予測に役立つならば構造的文法規則も考慮されると考える方が自然であろう。ともあれ，人間と生成言語モデルの言語規則は疑いなく異なる。

3つ目の論点は，言語獲得についてである。私たちは言語を自在に使用するが，その「言語能力」をどのように獲得したのか説明も理解もできない。娘が生まれてきたとき，どうやって言語を教えればよいのか途方に暮れた。言語の存在も規則も全体像も知らない。つまり，言葉を使わず言葉を教える必要がある。そもそもそんなことが可能なのか。どうして子供は大人より新しい言語を獲得するのが上手なのか。どうして同様にネコに語りかけているのにネコは言語を獲得しないのか。この論理的な矛盾を私たちはすべて軽々と超えてきた。これが「人間の言語能力とは何か」というミステリーである。

言葉に依らず言葉を獲得するのであるから，言葉ではない「何か」，普通に考えると「発話された音」と「それ以外の感覚体験」の経験的共起性，つまり間接証拠の「恒常的連接」に依るしかなさそうである。しかし，赤ん坊に投げ掛けられた「パパだよ」は，目の前の人物を指すのか，動きを指すのか，表情を指すのか，あるいは，何らかの指示や問いかけなのか，無限の可能性がある。数学的には，これを十分な精度で同定するためには膨大な異なる組合せを実際に経験する必要がある。ところが，ホーンスティンが述べる通り，現実には赤ん坊は驚くほど少ない不完全な言語経験から母語を獲得する。また，経験から獲得されるなら，人や言語によって獲得に必要な時間にばらつきがありそうなものであるが，実際には言語獲得・言語発達は特定の言語に依存しない普遍的なパターンを呈する。

しかし，少ない不完全な経験下で言語獲得が起こる

＊1―高橋悠治・茂木健一郎：ATAK@ICC公開トーク「他者の痛みを感じられるか」2005年12月17日 https://www.nttticc.or.jp/ja/hive/artist-talk/20051217/

からといって，ほぼ出来上りのような「受け型」が経験に先んじて備わっているはずだ，とまでは短絡できない。それでは言語獲得の困難さを「経験に先んじる所与のもの」とみなすことでさりげなく迂回し，ミステリーの本質を先送りするだけである。言語は有限個の記号を組み合わせることで無限の表現を生成する。こうした「情報の組合せ構造(再帰性・構成性・階層性)」の認知は，言語に限らず外界からの感覚情報を読み解くのに不可欠な共通の土台である。運動獲得や情景理解も実際に個人が経験した範囲だけでは困難である。赤ん坊は複雑な複合文を経験しないが，常に複合的で多義的な現実世界の文脈に晒される。また，言語能力が人間という種に付帯するならば，その成立に関わる「経験」とは種として晒されてきたあらゆる感覚情報も含む。つまり，各個人の言語体験だけではなく「進化過程で得たものを含むすべての感覚経験」を考えねばならない。GPT-4 は学習に画像タスクも含めることで精度が向上し，GPT-4v は与えられた画像の内容に関する文を生成できる。だが，生成言語モデルの事前学習で得た内部パラメタの値を「進化的に得た所与のもの」とみなし，少ない不完全な言語経験から(few-shot)言語獲得ができる，と解釈してみても意味がない。全く言語経験なし(zero-shot)でも高度な質問応答ができることになる。結局これでは，すべてはどこからを経験とみなすかという恣意的区分の問題に帰してしまう。

視点を変えれば，生成文法と生成言語モデルは驚くほど似ている。「組合せ構造」を解きほぐす(disentangle)情報処理は，生成文法だけでなく深層学習でもその中核にあるものである[2]。また，いずれも特定の言語に依らない共通一般規則を希求する。記号の関係性だけを正視し，意味や行為・感覚への接地を棚上げする。内部の仕組みは単純な計算に従うだけとし，人間の作為の関与を最小化しようとする。言語を計算とみなし，経済性や最適性に着目する。そして，両者共に帰納に何らかのバイアスを含める。生成言語モデルが基礎にしている機械学習は，データを帰納バイアスで内挿する方法である[3]。データの集積がいくらあっても帰納バイアス(仮説)なしでは説明も予測もない。事実の集積が科学でないことは，石の集積が家でないのと同じことである[4]。

もちろん両者には明確に異なる点がある。生成文法は「説明の論理」で，実際の文は生成しない。生成言語モデルは説明ではなく文生成のための統計的予測技術である。高い精度で予測できることは，理解や説明を何も意味しない[3]。顔認識や音声認識が実用レベルになっても，私たちのそうした能力が解明されるわけではない。ホーンスティンが「一般化問題」について的確に指摘する通り，機械学習にせよ何にせよ，有限個の観察を一般化する帰納バイアスは無限にあり，経験則はそのままでは「説明の論理」にはならない。ここで「論理」というとき，私が想定するのは「誰も夢にも疑おうとしないあの完全な厳密性」[4] である。普遍的再現性を有し反例 1 つで反証可能な論理形式と異なり，経験則は解釈次第で遡行的に法則に合うとも合わないとも言えてしまう。2, 4, □, 8 の□に 6 を見ることも 4 を見ること($n^3-7n^2+16n-8, n=1, 2, \cdots$)もできる[5]。経験論を揺さぶるのはいつの時代も「論理」であった[6]。数学や論理学に見るように，論理形式の演繹ならば，有限個の規則で無限の実例を説明できる潜在性がある。一方，帰納と演繹の間には，有限と無限の間の永遠に埋まることのない深淵が広がっている。組合せの数は指数的爆発を伴うため，1 googol(10^{100})が宇宙に存在する原子の数より多いことを考えれば，「論理」を経験則で十分に近似することは現実的には不可能である。そして，私たちに説明の論理が必要なのは，世界の法則がそうあるからではなく，認知キャパシティに収まる捨象された範囲でしか「理解」を体験できない，私たち自身の問題である。

では，生成言語モデルが見せるあの言語能力は何なのか。意味を全く考えないのに，対話，要約，翻訳，質問応答，言い換え，文体変換，テキスト平易化，文法誤り訂正など，多様な言語的実践を実用に足るレベルでやってのけることなど可能なのか。最後にこれを考えるにあたり，冒頭の「痛み」の話に戻ろう。

痛みを言語で正しく表出できるかという問いは，ウィトゲンシュタインが『哲学探究』で繰り返す大変有名な例である。哲学的問題は言語の働きの誤解から生じると考え，日常の言語実践を何より重視したかれにとって，規則と論理や数学の基礎の問題は中心的関心

だったと考えられている[5,7,8]。「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奇妙な形で生成言語モデルに実装されている。生成言語モデルは意味を問わず，使用のみを見る。愚直すぎるほどに「大きな言語使用の総体」の圧縮と再現のみを追求する。文の最小構成単位であるトークンもバイト対符号化など使用にもとづいて決定される。各トークンを表す数値ベクトルの要素値も使用にもとづいて決定され，その「差異」のみが考慮される。一方で，「行為への接地」は決定的に欠落している。これは「言語の根源および原初的形態は反応であり，ここからのみ，より複雑な形態が成長する。言語は洗練化の過程である」[9]とする拠り所のはずである。代わりに，言語とは記号の並びを「使用」するときの社会的な取り決め(集団バイアス)である，という一面だけを統計モデルで掬い取る。これは異なる種類の言語ゲームなのである。

生成言語モデルは，できるだけ有用な文を生成するための工学的技術に過ぎない。外部プログラムを適宜呼び出し実行することはできても，生成文自体は論理的一貫性を欠き，銀行の基幹システムや自動外科手術を任せる類のものではない。事実，ChatGPT を有効活用するには，出力された文が自分の求めるものであるか十分判断できる必要がある。自力で答えを見つけるのはかなりの手間だが，答えがあっているかどうかは素早く判断できる場合のみ真価を発揮する。入力文は単に求める答えを引き出す「呪文」であり，その意味で「プロンプト」と呼ばれる。もしあなたが ChatGPT と対話して，「やはり人間とは違う」とか「大したことない」とか，暗に期待していたものと違うと感じるとしたら，日常言語との表面的類似性のせいで，この「呪文の体系」の言語ゲームを誤解している。世の中にあふれるプロンプト職人たちの力作を眺めてみるとよい。入力文で画像を生成するにじジャーニー*2の Discord を覗いてみるとよい。そこではプロンプトは現に「呪文」と呼ばれている。あなたは呪文の唱え方を，つまり「使用」を間違えているだけなのである。

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。」[10]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。

文献

1—L. Berglund et al.: “The Reversal Curse: LLMs trained on “A is B” fail to learn “B is A”.” arXiv:2309.12288(2023)
2—Y. Bengio et al.: IEEE Trans. Pattern Anal. Mach. Intell., **35**(8), 1798(2013)
3—瀧川一学: 自動車技術，**77**(10), 26(2023)
4—ポアンカレ(河野伊三郎訳):『科学と仮説』(岩波文庫)．岩波書店(1938)
5—水本正晴:『ウィトゲンシュタイン vs. チューリング——計算，AI，ロボットの哲学』．勁草書房(2012)
6—瀧川一学: 人工知能，**34**(5), 603(2019)
7—ジャック・ブーヴレス(中川大，村上友一訳):『規則の力——ウィトゲンシュタインと必然性の発明』．法政大学出版局(2014)
8—岡本賢吾:『ウィトゲンシュタイン——没後 60 年，ほんとうに哲学するために』所収，河出書房新社(2011)pp. 57–78
9—L. Wittgenstein: *Culture and Value*(revised edition)．Wiley-Blackwell(1998); L. ヴィトゲンシュタイン(丸田健訳):『哲学の歴史 第 11 巻 論理・数学・言語 20 世紀 II』所収．飯田隆編，中央公論新社(2007)p. 404
10—L. ヴィトゲンシュタイン(丘沢静也訳):『反哲学的断章(新版)』．青土社(1995)p. 37

＊2—https://nijijourney.com/ja/

解説
折田奈甫

今回の記事は，本連載のためにホーンスティンが 2023 年 7 月頃に書き下ろしたものである。2022 年後半から広く認知され始めた ChatGPT に代表される生成 AI の背後にある大規模言語モデルが人間の言語能力のモデルとして妥当ではないと主張している。

大規模言語モデルは，ニューラルネットワークという機械学習モデルを用い，大量の文章データから学習した言語モデル——次の単語を予測するモデル——である。ニューラルネットワークを使うことで意味的に似ている表現をとらえることができるようになり，学習データには現れなかった文脈であっても似ている文脈から次の単語を予測できるようになった。学習データは，日本語で

概算すると数千億文字から数兆文字のテキストデータに相当する(文献1のp. 72)。周囲の文章・文脈から次の単語が何かを予測するという自己教師あり学習*1によって，まるで人間が理解して書いているかのように応答する生成AIが実現した。日本語で書かれたわかりやすい入門書が出ているので詳細はこれらを参照されたい[1,2]。

ホーンスティンの論点は以下の3点にまとめられる。本解説では，岡野原氏と瀧川氏による指定討論をふまえて順に補足と議論をしていきたい。

(1) ChatGPTは人間のように代名詞を解釈しない。
(2) ChatGPTは人間ならありえない文法規則を「獲得」している。
(3) ChatGPTの学習データと人間の子どものそれは，質・量ともに大きく異なる。

まず，プロンプトを通してChatGPTに文の容認性判断をさせるというリッツィのデモストレーションは，両氏が指摘するように結果の解釈に注意が必要で，ChatGPTが人間のように文を解釈していないという根拠にはならないだろう。補足すると，言語モデルの容認性判断と人間の容認性判断を心理言語学的に比較評価する方法はすでにいくつか提案されている[3]。同様に，ChatGPTが人間ならありえない文法規則を「獲得」しているという指摘も，両氏の指定討論の通りモデルを正当に評価しているとは言い難い。

3点目の論点は，両氏が同意するように極めて妥当と考える。人間の子どもの言語インプットと大規模言語モデルの学習データが質・量ともに大きく乖離している点については，さまざまな専門家が指摘している[4]。ここでは発達心理学者マイケル・フランクによる，大規模言語モデルと人間の子どもの言語学習の違いに関する論考を簡潔に紹介したい[5]。フランクは，大規模言語モデルのインプット量と，人間が20年間で読み聞きするだろう言語インプットの量を比較する。大規模言語モデルの場合はモデルによって数千億から数兆トークンである一方，人間が生まれてから20年間で受けるだろう言語インプットは，識字ができて4億語，識字の助けを抜きにして下限で3000万語と推定している。ちなみに，人間の子どもと同程度のインプットにスケールダウンしたデータで大規模言語モデルを学習させるという取り組みは始まっているが[6]，どの程度見込みがあるのかは現時点ではわからない。フランクは，インプット量の他にも，人間の子どもは言語以外のマルチモーダルな情報(たとえば視覚，触覚，嗅覚などの情報)や，他の人間とのインタラクションを通して言語を学習すること，そして評価方法が大規模言語モデルと人間の子ども対象の実験とで大きく異なることも指摘している(子どもが対象の実験は子どもにわかりやすいように作られている)。

また，この短い論考でフランクは，人間には進化の過程で得られた生得的な知識・バイアスが備わっているから大規模言語モデルのような膨大な量の言語データを用いなくても学習できるのだとも推察している。進化については瀧川氏の指定討論でも触れられており，経験に先んじる生得的知識を仮定することは言語獲得の問題を先送りにしているだけで，生得的知識を問うならば「個人の言語体験」だけでなく進化の過程で得たものを含むすべての感覚経験を考えねばならないと指摘している。実際のところ，言語学者は共時的な言語獲得と言語進化の両方を最も重要な問題としてとらえ取り組んできた。言語獲得におけるマルチモーダル情報の役割についての研究も多い。言語以外の感覚情報が豊富であれば生得的言語知識は必要ないという単純な話にはならず，言語以外の要因を探ることで生得的言語知識の有様が見えてくるような面白さがある[7]。生得的な言語知識がどのようなものかは理論や仮説に依存し，重要な研究テーマであり続けている。一方で，生得的な言語知識を仮定せず，データの統計的な性質と言語に特化しない一般的な学習則やバイアスで言語の知識を全て学習できるという経験主義的な仮説はわかりやすい。生得的言語知識を仮定するのは，言語獲得の問題を先送りにしたいからではない。これまで何十年とかけて多くの言語学者が言語獲得の過程と最終的に到達する言語知識を研究してきた。この積み重ねを見ると手放すにはまだ早いと考えざるをえない，そういう仮説なのだと思っている。観察データの統計的性質から学ぶ側面はある。しかしそれだけでは説明できないことが多い。

次に，大規模言語モデルが人間の言語能力や言語処理の理解に役立つ可能性について考えてみたい。大規模言語モデルは，次にくる単語を予測できるように学習することでさまざまな言語処理を高精度で実現している。岡野原氏が著書で推測しているように(文献1のp. 69)，人間も予測による学習で言語を理解できるようになるのだろうか。単語の予測学習をするには何らかの先行知識が必要であり(何もない状態から予測は生まれない)，卵が先か鶏が先かの循環的問題を呈している。筆者の専門の

＊1―「教師あり」とは正解データを用いた学習を指し，「自己」とは正解を人手で用意するのではなく，1つのデータから穴埋め問題と正解の両方を自動的に生成できることを指す。一般の「教師あり学習」に比べて，「自己教師あり学習」はコストをかけずに大量の訓練データを用意できる。

心理言語学という分野では，子どもが次に出てくる言葉を予測できるか，予測と実際に出現した言葉との差から生じるエラーシグナルをどのように言語の学習に用いるか，というような問題をめぐってさまざまな実験が行われてきた。これらの研究を大雑把にまとめると，子どもは，音韻レベルの精密な予測(例：“an”の次には“ball”ではなく“ice cream”がくる)は難しいようだが，意味や構造などのレベルでは次に出てくる言葉をある程度予測できる。しかし，予測エラーが言語の学習にどのように影響し用いられるのかについてはわかっていないことが多く，結果もさまざまで一致しない。たとえば，予期しない状況で出現する新しい言葉の方が学習されやすいと報告する実験がある一方で，予測しやすい方が新しい言葉を学びやすいこと，予測エラーの強度が新しく学ぶ言葉の符号化に影響しないことを示す実験もある[8]。このように，人間の子どもの言語獲得における予測の役割についてはわかっていないことが多い。予測モデルのフィードバックが，どういう言語知識を獲得するためにどのように使われるのか，このような問いに対して明確で新しい知見が得られるのならば，ニューラル言語モデルが人間の言語獲得の研究に役立つ可能性があるかもしれない。

次に考えたいのは，モデルの認知科学的・言語学的妥当性だ。岡野原氏の指定討論によると，タスクによって構成性や再帰性をデータから学習し問題を解くことができるという。モデルが既存の文法理論に対応する内部状態をもつことを示す実験も紹介されている。しかし，次の単語を文脈から予測することで学習する大規模言語モデルは，一見して特別な事前知識を必要としない単純な仕組みに思えるが，実際のところは人間にはまだよく理解できていないバイアスを組み込んでいると考える方が経験的にも妥当だと示唆する研究もある(文献9のpp. 6-7)。学習データが質・量ともに人間の受けるインプットと比較して異質であることもふまえると，これらのモデルが獲得した規則なり状態なりが人間の言語処理や言語獲得に対して何を示唆するのかはまだ明確ではない。そもそも，これらのモデルはニューラルネットワークである。ニューラルネットワークは，脳の神経回路の基本的な特徴を模してはいるが，脳のモデルではない。深層学習が誕生する以前はコネクショニストモデルとも呼ばれ，認知科学の研究にも用いられてきた。しかし，1980年代のコネクショニスト批判で指摘された問題点は今も手つかずのままだ(過去のコネクショニスト批判については次田瞬氏による入門書[2]でわかりやすくまとめられている)。神経科学者のデビッド・マーは，情報処理システムとして脳を理解するために，計算理論(計算の目的，システムの入出力，制約などを明らかにする)，表現・アルゴリズム(入出力の表現とこれらを変換するアルゴリズム)，実装(上のレベルの計算を脳でどのように実現するか)の3レベルに分けて考えることを提案した[10]。マーの3レベルが絶対というわけではないが，言語処理・言語獲得の問題を考えるとき，これに代わる見通しのよい切り分けを筆者は寡聞にして知らない。1980年代のコネクショニスト批判の頃から，ニューラルネットワークは実装レベルとして根拠がないことが指摘されており，認知科学としての言語学におけるニューラルネットワークモデルは計算理論レベルの問題を扱っていることが多い。他の計算モデルでも構成性や再帰性などは表現できるし，他のモデルの方が圧倒的に解釈しやすいにもかかわらず，計算理論レベルの問題で，なぜあえてニューラルネットワークを用いる必要があるのか，筆者はずっと疑問だ。とはいえ，すべての言語学者がこのように考えているわけではない。ニューラルネットワークが生成文法のような言語理論においても有用だと考える言語学者はいる[11, 12]。

瀧川氏が述べるように，生成文法は説明を求める。生成文法は言語獲得を説明できる言語理論であろうとしている。筆者が人間の言語獲得・言語処理の文脈で出てくるニューラルネットワークに懐疑的なのは，結局のところ「説明」がないと思うからなのだろう。これは瀧川氏が言うように「認知キャパシティに収まる捨象された範囲でしか「理解」を体験できない，私たち自身の問題」なのかもしれない。しかし，人間の認知能力の限界を認めつつも，泥臭く地道に理解と説明を求めていくしかないと思っている。今回の技術はこれまでの――仮説，予測，実験――という科学の方法を問うているのかもしれない。因果関係や説明を求めるのをやめ，深層学習によって「発見」されたパターンや相関をもとにして，科学としての言語学は先に進めるのか。筆者は人間が「言語」を用いて考えてきた言語理論と実験の積み重ねの上でまだまだやることがあると思っている。

文献

1―岡野原大輔:『大規模言語モデルは新たな知能か』. 岩波書店(2023)
2―次田瞬:『意味がわかるAI入門――自然言語処理をめぐる哲学の挑戦』. 筑摩書房(2023)
3―E. G. Wilcox et al.: Linguistic Inquiry(2023) https://doi.org/10.1162/ling_a_00491
4―J. Kodner et al.: lingbuzz/007485(2023)
5―M. C. Frank: Trends in Cognitive Sciences, **27**(11), 990(2023)
6―BabyLM Challenge: https://babylm.github.io/
7―L. R. Gleitman et al.: Annual Review of Linguistics, **5**, 1(2019)
8―C. Gambi et al.: Cognition, **211**, 104650(2021)

9—M. Baroni: in *Algebraic Structures in Natural Language*. S. Lappin and J.-P. Bernardy eds., CRC Press(2022) pp. 5–22
10—D. Marr: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman(1982); デビッド・マー(乾敏郎・安藤広志訳):『ビジョン――視覚の計算理論と脳内表現』産業図書(1987)
11—J. Pater: Language, **95**(1), e41(2019)
12—T. Linzen & M. Baroni: Annual Review of Linguistics, **7**, 195(2021)

タイトル画像クレジット:vladystock/123RF

＊1 月号掲載予定の連載第 3 回は言語進化を取り上げ,
比較認知発達の専門家による指定討論を予定しています。

ノバート・ホーンスティン
Norbert Hornstein
メリーランド大学言語学科名誉教授(生成文法・統語論)

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)

藤井友比呂 ふじい　ともひろ
横浜国立大学大学院環境情報研究院教授(統語論)

小野 創 おの　はじめ
津田塾大学学芸学部教授(文処理・心理言語学)

岡野原大輔 おかのはら　だいすけ
Preferred Networks 代表取締役最高研究責任者

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)

AI has never been a modest field
Norbert Hornstein(July 2023, 書き下ろし)