言語研究における科学的理解と言語モデル

鈴木陽登 [1]　菅原朔 [2]
[1] 慶應義塾大学大学院　[2] 国立情報学研究所
s17599@keio.jp　saku@nii.ac.jp

概要

近年の大規模言語モデルの急速な発展を受けて、言語研究の領域では言語モデルの言語学的・科学的な意義について活発な議論が行われている。しかし、そのような議論において、各々の研究者が（自身がすでに受け入れている）特定の理論的前提に基づいて必ずしもフェアとは言えない主張を展開している場面がしばしば見受けられる。このような背景から、本稿では、言語モデルの科学的理解への貢献を評価する際に考慮すべき要件を、科学哲学における**文脈主義**の立場を援用しつつ明確化することを目的とする。そして、文脈主義の枠組みに基づき、言語研究における言語モデルの有用性を評価するためには、(1) 理解の対象は何か、(2) 使用する理論は何か、(3) 理論の使用者は誰かの 3 点を考慮する必要があると主張する。

1　はじめに

近年、GPT をはじめとする大規模言語モデルの急速な発展により、これらのモデルが示す高度な言語運用能力を研究することによって、ヒトの言語能力の本質に関する重要な示唆をもたらす可能性が指摘されている。

しかし、言語モデルがヒトの言語能力についての科学的理解に実際に貢献しうるのかという点については、研究者の間で見解が大きく分かれている。このような対立の背景には、そもそも言語モデルが科学的理解をもたらすために必要な条件とは何かという点について、明確な合意が形成されていないという現状がある。その結果、各研究者が自身の理論的前提に基づいて恣意的に条件を設定してしまい、建設的な議論の妨げとなっているように見受けられる。

この問題に対して、本稿では近年の科学哲学における科学的理解の本性をめぐる議論が有益な示唆を与えうると考える。本稿の目的は、言語モデルの探求が科学的理解に資するか否かを論じる際に明確にしておくべき要件を、科学的理解についての文脈主義という立場に基づいて設定することにある。

2　言語モデルの意義をめぐる論争

言語モデルを言語研究に用いることの意義については、言語学者や心理学者の間で既に意見の対立がある。一方の極には、言語モデルはその工学的な実用性を超えて、ヒトの言語能力についての真なる理論を構築するという理論的言語学の目的に資することはないとする論者がいる [1, 2, 3]。もう一方の極には、言語モデルは既存の言語理論よりも優れた、あるいは最良の言語理論として扱われるべきだと主張する論者が存在する [4, 5]。

本稿では、言語研究における言語モデルの意義に懐疑的な論者と肯定的な論者の論争にはある問題が存在すると考える。それは、両者の主張が、それぞれの理論的前提に強く依存しているように見えるという点である。例えば、懐疑的な論者は、言語が合成性や言語能力と言語運用（competence–performance）の区別、モジュール性などを併せ持つシステマティックな対象であると見なし、言語学の目的はそのような言語に対する明示的な説明理論を構築することであると考える傾向がある [1, 6, 7]。一方で言語モデルに肯定的な論者は、言語を創発的な複雑系として捉え、モデルの予測の精度を重視する科学観に立脚する傾向がある (cf. Piantadosi 2023[4]: 26-7)。しかしながら、言語モデルが科学的理解に資するか否かを論じるにあたって、このような理論依存的な観点を採用してしまうことは、モデルや理論の科学的地位を必要以上に制限してしまう可能性がある。例えば、言語モデルは言語能力と言語運用の区別やモジュール性、合成性などの特徴を示さないため言語の科学的理論とは見なされないとする主張は（cf. Fox & Katzir 2024 [8]）は、それらの性質を必ずしも受け入れるわけではない言語研究上のアプローチ（例えば用法基盤モデル

に基づく構文文法）もまた科学的理論ではないことにしてしまう [9]。

このような問題が示唆することは、一口に「言語モデルは（ヒトの言語能力の）科学的理解に資する / 資することはない」と言ったとしても、そのような言明は実際は高度に理論負荷的な（すなわち、前提となる理論に大きく影響を受ける）ものであり、その適切性の評価もまた、特定の理論や文脈に位置付けた上でなされる必要があるということである。さもなければ言語モデルの科学的意義をめぐる議論は、異なる理論的前提に基づいたまま平行線をたどるだけに終始してしまう危険性がある。以下ではこの点をさらに詳細に論じるために、科学哲学における科学的理解をめぐる議論、とりわけ Henk de Regt が提唱した科学的理解の文脈主義と呼ばれる立場を導入する。

3　科学的理解の文脈主義

科学的理解の本性をめぐる議論は、近年の科学哲学において中心的な研究領域の一つとなっている[1)]。科学的理解の領域で問題とされている「理解」とは、ある現象がなぜ生じたのかについての理解（**why-理解**）である。この種の理解についての有力な学説の一つとして、De Regt による文脈主義が挙げられる [11]。De Regt は理解を主体のある種の能力として捉える「能力説」の立場から、科学的理解の条件を以下のように定式化している。

> 主体 S が理論 T に基づいて現象 P の適切な説明を構築できるとき、S は P を理解している。(De Regt 2019 [12]: 67)

より具体的には、De Regt は**現象の科学的理解の基準（Criterion for Understanding Phenomena; CUP）**を提示している。

> CUP: ある現象 P が科学的に理解されるのは、P に関する説明が**可解な（intelligible）理論 T に基づいており**、かつその説明が内的整合性や経験的十全性といった基本的な認識的価値を満たしているときである。（De Regt 2017 [11]: 92）

ここで鍵となる、理論の**「可解性」（intelligibility）**という概念は、直観的には科学者がある理論を効果的に説明やモデル構築のために使用できる程度として理解される。より正確には、理論の可解性とはある科学理論の性質の集合に科学者が帰属する価値として定義される。具体的には単純性や整合性、スコープの広さなどの一般的な理論的美徳のほか、因果性、可視性、数学的抽象性などが可解性に資する理論の性質として挙げられる。そして、これらの性質は科学者による理論の使用を促進する役割を果たす。ここで重要なのは、どの理論的性質が可解性に寄与するかは、理論を使用する個々の科学者やその属する科学者集団が持つスキルや背景知識に依存するという点である。すなわち、同じ美徳を持つ理論であっても、時代や文脈に応じて科学者にとって可解であったりなかったりすることがある。

では具体的にどのような規準を満たした理論が可解とされるのだろうか。理論の可解性が科学者（集団）のスキルや知識といった文脈に依存するということは、可解性という概念が単に好みの問題であるということにはならない。例えば De Regt は、**理論の可解性の基準（Criterion for the Intelligibility of Theories; CIT）**の一つとして、「（ある文脈 C において）厳密な計算なしに理論の定性的な帰結を認識できること」を挙げている（De Regt 2017 [11]: 102）。

ここで「理論の定性的な帰結を認識できること」とはどのようなことだろうか。De Regt は気体分子運動論の事例を持ち出してこれを説明している (De Regt 2017 [11]: 104-5)。分子運動論において、ある容器の圧力は、気体分子が容器の壁に衝突することで発生する（壁を押す）力の総体として捉えられる。ここで気体分子運動論における「熱は分子の運動である」という基本原理を前提すると、圧力と温度の関係を定性的に導出することができる。すなわち、ある大きさの容器に熱を加えると、気体分子の運動速度が速くなり、気体分子が容器の壁に衝突する頻度が上昇することで容器全体の圧力が増す。同様に、容器の体積が小さくなると（気体分子が容器の壁に衝突する頻度が上昇するため）容器内の気体の温度は上昇する。この推論において、詳細な計算は存在しない。De Regt によれば、むしろそのような計算は我々が既に持っている（理論に対する直観的な）理解によって動機づけられるのである（De Regt 2017 [11]: 105）。

以上で概観した文脈主義的アプローチは、科学的理解を論じる際に特定しておかなければならない事柄を明確にするという点で本稿の目的にとって有益である。CUP によると、現象 P が理解できるかどう

1)　現在の科学的理解論の議論状況について日本語で読める文献として [10] が挙げられる。

かは、理論 T の可解性（および整合性・経験的十全性）に依存し、さらに可解性という概念は理論の使用者のスキルや背景知識といった文脈に依存するのであった。したがって、科学的理解を論じるには、少なくとも (1) そもそも理解したい現象は何か、(2) 使用する理論は何か、(3) 理論の使用者は誰か（どのようなスキルや知識を持った主体か）という 3 点が明らかになっていなければならない。

4　言語モデルと文脈主義

本節では、言語研究における言語モデルの役割をめぐる議論を、(1) そもそも理解したい現象は何か、(2) 使用する理論は何か、(3) 理論の使用者は誰かという 3 つの観点から明確化するための指針を提供することを試みる。

4.1　理解したい現象は何か

まず、言語モデルを用いて理解しようとする現象が何であるかを明確にする必要がある。言語学における研究対象が言語であるといっても、言語には様々な側面・階層が存在する。例えば、音韻論・音声学、形態論、統語論、意味論、語用論という、現在広く受け入れられている言語学の分野は、言語が有する階層に基づいて区分されていると言えよう。さらに、言語研究者たちは上記の領域に加えて、言語獲得/習得のメカニズム、文処理（文の産出・理解）のプロセス、言語の進化といった側面にも注目している。また、言語の特定の側面を切り出したとしても、それをどの抽象度で明らかにするのかという問題が出てくる。例えば同じ統語論に属する研究であっても、関係節構文や wh 疑問文などの具体的な言語現象を詳細に記述する研究や、データを用いて理論を検証・洗練させる研究、既存の統語論研究で提案されてきた一般則を統一しようとする研究など、その抽象度に応じて様々なバリエーションがある（cf. ホーンスティン 2024 [13]）。

したがって、言語研究における言語モデルの役割を論じる際には、まず言語のどの側面・階層に注目するか、そしてその側面や階層をどの抽象度で研究するのかという点が明らかになっていなければならない。例えば Ambridge と Blything は言語研究における言語モデルの役割を肯定的に評価しているが、その射程は動詞の項構造という具体的な統語現象という領域においてである [5]。

4.2　使用する理論は何か

理解したい言語現象は何かが特定されると、次はそれに対する理論は何かという問題が浮き上がる。そして、とりわけ言語研究における言語モデルの意義をめぐる論争で問題となるのは、言語モデルそれ自体をある種の言語理論として見なせるか、あるいは使えるか否かということである。

言語モデルのブラックボックス性を考慮すれば、この問いに対して否定的な態度を抱くのは自然である。実際、可解性の概念を導入した De Regt も、正確な予測を行う一方で、なぜその予測がされるのかが分からない神託のような理論を、可解でない理論の例として挙げている (De Regt 2017 [11]: 101)。科学的理解の文脈主義によれば、科学者にとって可解でない理論は、現象の理解にとって役に立たない。

しかし、Sullivan が指摘するように、言語モデルを含む機械学習モデルのブラックボックス性は程度の問題であり、必ずしも理解を阻害するとは限らない [14]。確かに機械学習モデルの実装レベルの複雑さや不透明性といった問題は残るものの、モデル製作者は完全に無知な状態でモデル構築をしているわけではない。モデル製作者は特定の統計的仮定や理論を使用してモデルの出力や予測を改善したり、訓練後のモデルのおおまかな挙動を把握したりすることができる。

実装レベルのブラックボックス性が理解を必ずしも阻害しない例として、Sullivan はシェリングの分居モデルを挙げている（Sullivan 2022 [14]: 117)。分居モデルは、様々な人種がなぜ互いに交じり合わずに分居するのかを説明する人間社会のモデルである。このモデルにおける主体は、隣の住民が一定程度自分と同じ人種であることを選好するという単純なアルゴリズムに従う。さて、シェリングのモデルをコンピューターで実装しようとしたとき、「隣人が一定程度同じ人種であることを好む」というアルゴリズムは様々なプログラミング言語や手法によって実装できる。しかし分居現象を理解したいモデルの使用者にとって、シェリングのモデルがどのように具体的に実装されているのかを知る必要はない。このように、実装レベルのブラックボックス性そのものが現象を説明したり理解したりする能力を損なうわけではない。

以上の考察から、言語モデルを理論としてみなせるか否か、あるいは理論として使えるか否かという

問いに答えるには、言語モデルのどのレベルのブラックボックス性に注目するかを明らかにしなければならない。あるレベルにおいてはブラックボックス性の程度が薄くなったり、あるいはそのブラックボックス性が無害である場合がある。

なお、言語モデルを理論の一種とみなす論者は、しばしばモデルの予測の精度と現象や理論の理解をトレードオフなものとして考える (cf. Ambridge & Blything 2024 [5]: 45)。しかし、理解と予測は必ずしもそのような関係にあるわけではない。De Regt によれば、科学者による予測の成功は、関連する理論を科学者が理解しているかに依存する一方で、理論を用いた予測の成功は、その理論が科学者にとって可解であるということを示している (De Regt 2017 [11]: 107-8)。このように、理解と予測は緊密に関係しており、トレードオフな関係にあるというよりは、むしろ一方が改善されることでもう一方も改善されるという関係にある（cf. Douglas 2009 [15]）。

4.3 理論の使用者は誰か

科学的理解の文脈主義による重要な指摘の一つとして、ある現象 P に対して理論 T が定まったとしても、P が実際に理解可能か否かは、理論を使用する主体の認知スキル、具体的には T を用いて P に対する適切な説明モデルを構築する能力に依存するということである。したがって、同じ理論 T が与えられたとしても、ある主体は P を適切に理解できる一方で、別の主体は P を理解できないという状況は十分に考えられる。科学的理解の成否を評価するためには理論の使用者がどのような認知スキルや背景知識を持っているのかについての考慮が不可欠である。

この考察を言語研究の文脈に適用すると、言語モデルの科学的理解への意義を評価する際には、モデルを使用する研究者の認知スキルや専門性を考慮する必要があるということがわかる。一般に、科学者が有するスキルセットは、その科学者が属する共同体（≒パラダイム）内で習得・評価される（cf. De Regt 2009 [16]）。そのため、言語研究者の認知スキルを考慮する際には、その研究者がどのパラダイムに属しているのかを確認するのが有益である。

これに伴って、特定のパラダイムにおける理論の可解性基準は何かという点も明らかにしなければならない。前述したように、De Regt は「厳密な計算なしに理論の定性的な帰結を認識できること」という理論の可解性基準を提案しているが、これはあくまでも物理学における可解性基準にすぎず、別の分野やパラダイムにおいても De Regt の基準が妥当するとは限らない。むしろ、特定の分野やパラダイムそれぞれに固有の認知スキルの体系が備わっていることを考慮すれば、分野やパラダイムごとに理論の可解性基準が異なると考えるのは自然である。したがって、言語研究における科学的理解を論じる際には、自分や相手がどの分野やパラダイムに属しており（生成文法、認知言語学、形式意味論など）、どのような認知スキルや背景知識を持っているのか、問題となる分野やパラダイムにおける理論の可解性基準は何かといった要素を定めなければならない。

5 まとめと今後の展望

本稿では、言語モデルが科学的理解に資するか否かを議論する際に明確にすべき条件を、De Regt の科学的理解の文脈主義に基づいて検討してきた。具体的には、(1) 理解したい現象は何か、(2) 使用する理論は何か、(3) 理論の使用者は誰かという 3 つの観点から、言語モデルの科学的意義を評価する際に考慮すべき要素を明らかにした。

De Regt の枠組みが示唆することは、言語モデルの科学的意義を一般的な形で論じることは困難であり、むしろ特定の文脈に即して評価する必要があるということである。言語モデルが、ある現象についての理論やモデルとして有用であるか否かは、研究者がどの抽象度で現象と向き合うか、その研究者（集団）の持つスキルや背景知識、所属するパラダイムにおける理論の可解性基準は何かといった要素を特定することによって、はじめて評価できる。

今後の課題としては、主に各言語研究のパラダイムにおける理論の可解性基準を特定することが挙げられる。そうすることによって、言語モデルがそれぞれのパラダイムにおいてどのように科学的理解に貢献しうるのかを、より具体的に評価することが可能となると考える。

本稿で提示した枠組みは、言語モデルの科学的意義をめぐる議論を整理する一つの視座を提供するものである。この枠組みが、言語研究における言語モデルの可能性と限界についてのよりフェアな議論の一助になれば幸いである。

謝辞

本研究は JST 創発的研究支援事業 JPMJFR232R の支援を受けたものです。

参考文献

[1] Gabe Dupre. (what) can deep learning contribute to theoretical linguistics? **Minds and Machines**, Vol. 31, No. 4, pp. 617–635, 2021.

[2] Noam Chomsky, Ian Roberts, and Jeffrey Watumull. Noam Chomsky: The false promise of ChatGPT. **The New York Times**, 05 2023.

[3] ノバート・ホーンスティン. 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. 科学, Vol. 93, No. 12, 12 2023.

[4] Steven T. Piantadosi. Modern language models refute Chomsky's approach to language. In Edward Gibson and Moshe Poliak, editors, **From fieldwork to linguistic theory: A tribute to Dan Everett**, Empirically Oriented Theoretical Morphology and Syntax 15, pp. 353–414. Language Science Press, 2024.

[5] Ben Ambridge and Liam Blything. Large language models are better than theoretical linguists at theoretical linguistics. **Theoretical Linguistics**, Vol. 50, No. 1-2, pp. 33–48, 2024.

[6] Jordan Kodner, Sarah Payne, and Jeffrey Heinz. Why linguistics will thrive in the 21st century: A reply to piantadosi (2023), 2023.

[7] Roni Katzir. Why large language models are poor theories of human linguistic cognition: A reply to piantadosi. **Biolinguistics**, Vol. 17, , 12 2023.

[8] Danny Fox and Roni Katzir. Large language models and theoretical linguistics. **Theoretical Linguistics**, Vol. 50, No. 1-2, pp. 71–76, 2024.

[9] Stefan Müller. Large language models: The best linguistic theory, a wrong linguistic theory, or no linguistic theory at all? **Zeitschrift für Sprachwissenschaft**, 2024.

[10] 小林佑太. 説明的理解論の現在——把握・知識・理解. フィルカル, Vol. 6, No. 2, pp. 180–205, 8 2021.

[11] Henk W. de Regt. **Understanding Scientific Understanding**. OUP USA, New York, 2017.

[12] Henk W de Regt and Christoph Baumberger. What is scientific understanding and how can it be achieved? **What Is Scientific Knowledge?**, pp. 66–81, 2019.

[13] ノバート・ホーンスティン, 折田奈甫, 藤井友比呂, 小野創, 窪田悠介. 人間の言語能力とは何か-生成文法からの問い〈6〉真に理論的な統語論の研究はどれぐらいあるのか. 科学, Vol. 94, No. 6, pp. 556–563, 05 2024.

[14] Emily Sullivan. Understanding from machine learning models. **British Journal for the Philosophy of Science**, Vol. 73, No. 1, pp. 109–133, 2022.

[15] Heather E. Douglas. Reintroducing prediction to explanation. **Philosophy of Science**, Vol. 76, No. 4, pp. 444–463, 2009.

[16] Henk W. de Regt. **Scientific Understanding: Philosophical Perspectives**. University of Pittsburgh Press, 2009.