

大脳新皮質の原理理解に向けた計算論的領野比較

Computational cross-areal comparison for understanding of neocortical computational principles

寺島 裕貴

Hiroki Terashima

NTT コミュニケーション科学基礎研究所

NTT Communication Science Laboratories

Computational neuroscientists have long been interested in deciphering computational principles of the neocortex, the locus of intelligence. However, despite of the advances of neurophysiology, we have no clear idea yet on how to deduce functional universals by comparing multiple functional areas. In the present study, we show that our primary difficulty resides in the incomparability of functions, and that Linguistics has conquered it using a model-driven approach. As Generative Grammar tested candidate English models in other European languages, computational neuroscientists have started to test visual cortex models in the auditory system, which enables them to compare the two systems from non-trivial viewpoints. To elucidate “universal functions” of the neocortex, we should further this kind of computational cross-areal comparisons.

1. 大脳新皮質の計算原理探求

大脳新皮質の持つ二面性は、古くから研究者を惹きつけてきた [8]。およそすべての知能にかかわる機能を支える新皮質は多くの領野にわかれ、それぞれが多様な機能を呈している。その一方でそれら領野は基本的な解剖学的構造を共有しているため [9, 10]、何らかの共通計算原理の存在が示唆されてきた。この普遍性を機能の多様性の中にどうやって見出すのが計算神経科学の大きな挑戦である (図 1)。

近年の技術進歩を受け、大脳新皮質の各領野における神経生理学は大きな進歩を遂げてきた。特に、視覚を司る視覚野の機能解明の進展は著しい。にもかかわらず、新皮質全体の計算原理に関わる理解はあまり進んでいないのが実状である。本稿では、この探求の何が難しいのか、困難を解決するためのヒントはどこにあるのか、そしてそのヒントをもとに今後の大脳新皮質研究がどう進むべきかを議論する。

なお、本稿は基本的なアイデアを著者の博士論文 [23] を初出とし、その後の考察を反映したものである。

2. 機能の比較困難性という問題

新皮質の領野間で機能の普遍性を見出す試みはあまり成功していないが、領野間で普遍性を見出す試みの全てがうまくいっていないというわけではない。古典的な成功例は、Brodmann による大脳新皮質の細胞組織学的な領野比較に見ることができる [5]。彼はまず、新皮質全体から網羅的に細胞の染色像を収集した。その上でそれらを丹念に比較することで、新皮質が多くの領野にわかれることを示し、それは後に機能的な区分とも概ね一致することがわかった。Brodmann は徹底的な比較で差異を明らかにした上で、それでもなおそれら領野は基本的な解剖学的構造を共有していると主張した。

計算原理研究と Brodmann の成果との違いは何だろうか。彼の研究法は、データが遍く収集可能であること、さらにそれらが直接比較であることを前提としていた。しかし機能や計算

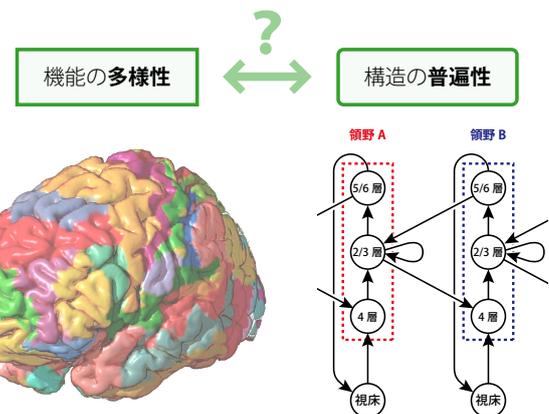


図 1: 大脳新皮質の多様性と普遍性。大脳新皮質の領野は多様な機能を呈すが (左)、何らかの計算原理の存在が構造の普遍性 (右) から示唆される。

といったレベルにおいては、これらの前提が成立しない。第一に、多数の新皮質領野のうち、機能が明確に議論できる領野の数は依然としてかなり限られている。比較すべきデータの入手可能性に領野間の差異が大きい。

第二に、こちらのほうがより深刻な問題だが、たとえ全領野の機能が明らかになったとしても、それら機能を直接比較することは難しい。わたしたちは各領野の機能を議論する際、それぞれのモダリティに特有の言葉を使って説明を試みてしまう。例えば視覚関連領野の機能を語るのに、視野の XY 座標、方位、方向、視差といった概念は欠かせない。もちろん、その行為自体には何の問題もないし、視覚機能を理解するのに最適な戦略であることは間違いない。しかし、いったん視覚の枠を超えて他のモダリティ関連領野との共通性を考えようとすると、視覚固有の概念を導入したことが足かせになってしまう。視覚野の方位 45 度ニューロンは、運動野の 45 度ニューロンと直接比較していいのだろうか。聴覚野の 45 度ニューロンとはいったい何だろうか。前頭葉の意思決定ニューロンと比較して、共通の機能を見つけられるだろうか。

この機能の比較困難性という問題は、機能の記述にモダリ

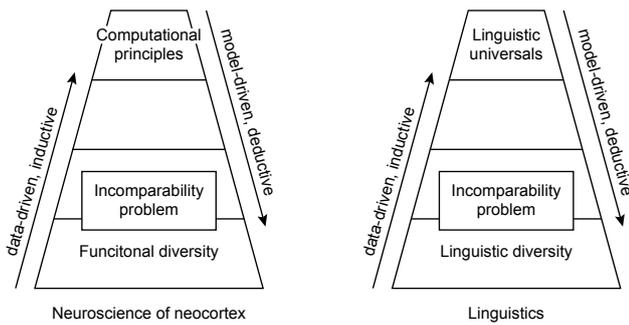


図 2: 普遍性への二つのアプローチ: ボトムアップとトップダウン. 言語学は, 言語の多様性の中に言語普遍性を見出そうとしてきた. その中で, ボトムアップな言語類型論ではなくトップダウン・モデル駆動型の生成文法を生み出した (右). 大脳新皮質の計算原理研究も類似の問題構造を持っている (左).

ティ固有の概念を使ってしまふ以上, 避けることはできない. 領野間で自明に比較可能なのは, Brodmann が用いた形態情報や遺伝子発現パターン [15, 18], スパイク統計 [21] といった物理量に限られ, 本質的には解釈に過ぎない機能のレベルでは比較可能性が担保されない. では, 機能の普遍性を追い求めるのは諦めるべきなのだろうか. 必ずしもそうではないことを示しているのが言語学の歴史である.

3. 言語学のモデル駆動型比較に学ぶ

3.1 言語類型論: ボトムアップ・データ駆動

いかにして多様性の中に普遍性を見出すか——という問題意識は, 多くの学問に共通している. しかし, 普遍性への強い確信を抱きつつも比較困難性に直面せざるを得ないという点で, 新皮質の計算原理研究は特に言語学と構造的に似通っていることを指摘したい. 言語学の長い歴史は, 新皮質研究に重要な示唆を与えてくれる.

どのようにすれば多様な言語の中に普遍的な特徴 (言語普遍性) を見出すことができるだろうか. この古くからの問いに対する古典的なアプローチは, 言語類型論 [25] である. 複数言語のデータをとにかく収集し, それらを丹念に比較すれば, 言語普遍性は自ずと明らかになるはずである. しかしその結果明らかになったのは, あらゆる言語に共通する特徴は——少なくとも人間が明らかに見いだせる範囲では——ほとんど無いということだった [14]. 真に共通していることといえば名詞の存在くらいのもので, 動詞さえ普遍的な品詞とは言えなかった. ボトムアップにデータ駆動で普遍性を見出そうとする試みは, 系統上遠い言語間や抽象的な概念間の比較で行き詰った (図 2 右).

3.2 生成文法: トップダウン・モデル駆動

この困難に対し打開策を与えたのが, Chomsky の生成文法 [6] である. 彼は普遍性を所与とし, 大胆にもまず単一のモデルを仮定した. その上でモデルを実言語にフィットさせ, そこで生まれる齟齬をフィードバックし, 理解可能性を維持できる範囲でモデルを改善していくというループを回すことを提案したのだ (図 2 右). その結果として, それまで言語間でまるで異なるとされていた概念が, 実は単一モデルのわずかなパラメータの違いで説明できることが示された. トップダウンでモデル駆動型の比較により, 理解可能性を担保したまま非自明な比較を提案できるようになり, 多言語の統一的理解が進んだ. 多くの場合はまず英語を基にモデルが構築され, それが他の欧

州の言語, さらに別の言語へと適用される中でモデルが改善されていくという形をとった.

普遍文法の歴史をたどると何度か大きな更新が行われており, どのようなモデルがより望ましいのかという考えの変遷が読み取れる. 中期の原理とパラメータのアプローチにおいて, 普遍文法は全言語に共通の原理と, 個別言語特異的な値をとるパラメータの組としてモデル化されていた. 当初は, 個別言語の振る舞いを説明するためにはパラメータを増やしていけばよいと思われていた. しかし, 実際に数多くの個別言語を記述しようとするパラメータの数が膨大になってしまう. 記述的妥当性は確保することができるが, 説明的妥当性が満たされないという本末転倒な結果に至ることが明らかになった [1]. Chomsky はその後, ミニマリストプログラム [7] を提案し, 最小モデルを模索するようになった.

この変遷は機械学習とのアナロジーとしても興味深い. パラメータを増やして記述的妥当性だけを確保するのは, 機械学習で言う過学習 (overfit) に対応する. やみくもにモデルの自由度を増やして訓練データを説明できるようになっても, テストデータの予測性能はむしろ悪化していつてしまう. ミニマリストプログラムは過学習を回避するための正則化つきの機械学習に相当するだろう.

4. 大脳新皮質のモデル駆動型領野間比較

4.1 トップダウン比較で最小モデルを目指す

言語学と大脳新皮質の計算原理研究は, 多様性の裏に普遍性の存在が信じられつつも比較困難性の壁に阻まれてきたという点で類似している. 言語学のほうがはるかに長い歴史を持ち, われわれはその歴史から学ぶことができる.

各領野の機能研究をまず行い, 明らかになった機能を比較することによって普遍性を追求しようとする従来のボトムアップアプローチは, 言語学でいう言語類型論に相当するだろう (図 2 左). 仮に全領野の機能が明らかになったとしても, それら機能を直接比較し, さらに共通性を抽出することはわたしたちには難しい. 全言語に共通の普遍性がほとんど見つからなかったように, 新皮質の全領野に共通の機能もほとんど見つからないことが予想される.

では, 新皮質領野の機能の普遍性についてトップダウンのアプローチは存在しなかったのだろうか. 影響力を持っている理論として, ベイズ推論と自由エネルギー原理 [11] がある. たしかに, これら理論は強力であり, 新皮質で見られる多くの現象を説明することができる. しかしそれで十分ではないこともまた, 生成文法の歴史によって示されている. いずれの理論もあまりに自由度が高く, どんな現象でも説明できようと思えばできてしまう. 現在の学界のステージは, そのような過学習の状態に留まっていると考えられる. 今後はこのような過学習を避け, いかにモデルのアーキテクチャを制限するかを考えていく必要がある. すなわち, ベイズ推論にせよ自由エネルギー原理にせよ, 具体的にどのようなベイズ推論なのか, あるいはどのような自由エネルギーなのか, 自由度を制限した上での説明能力を問うていくことになるだろう.

その際, 候補となるモデルは最もデータが豊富な領野, すなわち一次視覚野 (V1) の研究の中から生まれることが期待される. これは, 生成文法で英語ベースの理論がまず生まれてきたのと同様である. 英語で提案されたモデルが, まずは比較的近縁の言語でその普遍性をテストされ, さらに遠い言語で真の普遍性を試されてきた. 新皮質でも同様に, まず V1 で提案されたモデルが, まずは最も「近縁」[15, 18] の一次聴覚野 (A1)

でテストされ、いずれは全領野における普遍性が試されていくはずだ。次節で、このような計算論的領野比較の A1 における研究例に触れる。

なお、このようなトップダウン型のモデル化が脳の意識研究でも勃興しつつあるのは興味深い。意識研究は、個々の意識的体験を支える現象を積み上げていくだけではその本質が捉えられないという問題に直面してきた（ハードプロブレム）。近年になって、統合情報理論（IIT）は意識の現象論的特徴をトップダウンに公理として設定し、そこから導かれるモデルの更新を繰り返しはじめつつある [24, 19]。たとえば視覚的なクオリアと聴覚的なクオリアを統一的に説明することが可能になるのか、この種の手法の有効性がこれから試されることになる。

4.2 例：視覚野モデルから聴覚野へ

大脳新皮質の計算論的領野比較は、まず V1 で自由度が制限されたモデル候補を選び、次に A1 でテストするという流れが有望である。本節ではそのような試みの例として、地図構造のモデル化と深層学習を紹介する。

第一に地図構造について、V1 や A1 の皮質表面ではレチノトピー（視野再現地図）やトノトピー（周波数地図）の存在が知られていた。しかし、近年の技術的進歩により、微細スケールにおいては A1 の地図が V1 と比べはるかに乱れていることが明らかとなった [3, 20]。従来は地図構造の類似性が V1 と A1 の共通性の根拠の一つだったが、やはりこの二つの地図は異なる規則に支配されているのだろうか。

Terashima ら [22] は、この異なる地図構造が同一の学習則から自然と生じうることを指摘した。V1 の地図構造については、トポグラフィック独立成分分析（TICA）が自然画像の学習モデルとして過去に提案されていた [16]。彼らはこのモデルを自然画像のかわりに自然音に適用し、その学習結果が A1 に見られるような乱雑な地図となることを示した。この結果は、TICA という限定されたアーキテクチャが、少なくとも V1 と A1 に共通の学習の本質を捉えている可能性を示している。

さらに興味深いのは、このモデル適用によって V1 と A1 の間で非自明な比較が可能となった点にある。TICA は V1 の地図構造のみならず、非線形な複雑細胞のモデルでもあった。彼らは、このモデルを自然音に適用して得られた聴覚版のモデルにおいて、「A1 版複雑細胞」の反応を調べた。その結果、その非線形反応は、A1 でピッチ細胞として別に知られていた細胞 [4] とよく似ていることがわかった。V1 複雑細胞と A1 ピッチ細胞は、たしかに非線形な変換を行い不変性の抽出を行っていることができる。しかしこのモデル駆動の比較を行うまで、その非自明な相同性は一切指摘されていなかった。これはまさに、生成文法の導入で異なる言語の異なると思われる概念が結び付けられたことに相当しており、モデル駆動型の領野比較の有用性をよく示している。

第二に深層学習（deep learning）について、これは人工知能分野で近年注目を集めている技術だが、その発端が脳にヒントを得ていることから [12, 13]、脳とどのように対応するのも関心の的となっている。Yamins らは多層の畳み込みニューラルネットワーク（CNN）を画像識別タスクで訓練し、得られた中間層表現の階層性がヒト視覚系の階層構造とよく対応することを示した [26]。

Kell らはこの結果をもとに、多層 CNN に自然画像のかわりに自然音の識別タスクを課し、その結果得られる中間層表現の階層性がヒト聴覚野の階層構造と対応することを示しつつある [17]。この結果もまた、多層 CNN を自然なタスクで訓練するという限定されたアーキテクチャを検討しており、その結果として視覚野・聴覚野双方の階層構造の理解を与えてくれている。

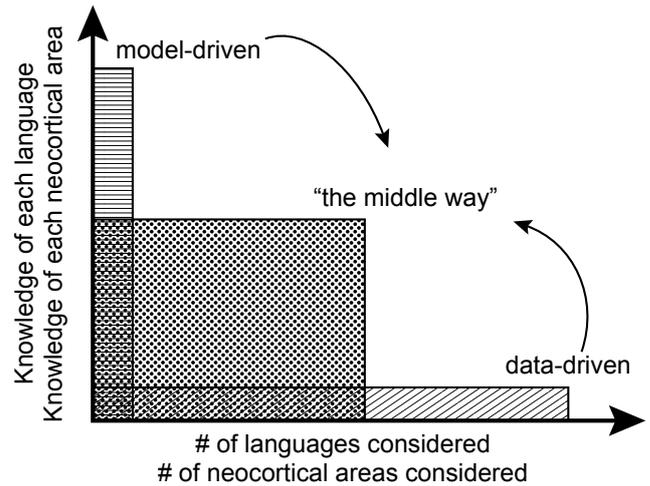


図 3: 未来：ボトムアップとトップダウンの「中庸」へ。文献 [2] をもとに改変。

る。中間層のさらなる解析により、非自明な比較が明らかになるのが期待される。

このように、V1 を起点とする計算論的領野比較は A1 をはじめとする聴覚野を舞台に始まりつつある。しかしその規模は決して大きいとは言えず、今後さらなる発展が期待される。ところで、言語学はさらに先の展開として何を指し示しているのだろうか。言語学では、生成文法と言語類型論の「中庸」を目指すべきではないかとの議論が生まれつつある [2] (図 3)。しかし両陣営には互いに相容れないところがあり、実際にはなかなか難しい。これは神経科学でも同じで、実験家と理論家は異なる行動原理を持っている。そのことは認めつつもいかに手を組んで共通の目的に対して前進できるのか、言語学の歴史に学ぶことで遠回りを回避していかなければならない。

5. まとめ

本稿では、大脳新皮質の計算原理解明がなぜ進まないのか、前進のためのヒントはどこにあるのか、どのような方向に進むべきかを議論した。

領野の多様性の中に普遍性を見出すという営みは、必然的に領野間の比較を必要とする。実際、大脳新皮質でも細胞形態や遺伝子発現パターンという物理量のレベルでは領野間比較が可能である。しかし、機能や計算というレベルでは比較できないというのが本質的な問題だ。

同様の問題に古くから直面していたのが言語学である。古典的な比較言語学は言語間のボトムアップな直接比較によって言語普遍性を抽出するのに挫折した。この問題に対して解決策を与えたのが、モデル駆動型の生成文法である。

言語学の歴史は、大脳新皮質研究の未来を暗示している。生成文法研究が英語から始まったように、大脳新皮質ではモデル仮説は視覚系から生まれ、他の領野との比較を通して普遍性が試されることになるだろう。その実例としてスパース性や深層学習といった概念が聴覚野へ適用され、ボトムアップな比較では辿りつけなかった相同性が指摘されつつある。言語学の歴史にならい、大脳新皮質の原理解明においてもモデル駆動型の領野間比較を推進していくべきだ。

参考文献

- [1] Baker MC. *The atoms of language: The mind's hidden rules of grammar*. Basic Books, 2001.
- [2] Baker MC. Formal generative typology. In Heine B and Narrog H, editors, *The Oxford Handbook of Linguistic Analysis*, pages 285–312. Oxford University Press, 2009.
- [3] Bandyopadhyay S, Shamma SA, and Kanold PO. Dichotomy of functional organization in the mouse auditory cortex. *Nature Neuroscience*, **13**(3):361–368, 2010.
- [4] Bendor D and Wang X. The neuronal representation of pitch in primate auditory cortex. *Nature*, **436**(7054):1161–1165, 2005.
- [5] Brodmann K. *Vergleichende Lokalisationslehre der Grosshirnrinde: in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Ja Barth, 1909.
- [6] Chomsky N. *Syntactic Structures*. Mouton, 1957.
- [7] Chomsky N. *The Minimalist Program*, volume 28 of *Current Studies in Linguistics Series*. MIT Press, 1995.
- [8] Creutzfeldt OD. Generality of the functional structure of the neocortex. *Naturwissenschaften*, **64**(10):507–517, 1977.
- [9] Douglas RJ and Martin KAC. Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, **27**:419–451, 2004.
- [10] Douglas RJ, Martin KAC, and Whitteridge D. A canonical microcircuit for neocortex. *Neural Computation*, **1**(4):480–488, 1989.
- [11] Friston K, Kilner J, and Harrison L. A free energy principle for the brain. *Journal of Physiology*, **100**(1):70–87, 2006.
- [12] Fukushima K. Neural network model for a mechanism of pattern recognition unaffected by shift in position — Neocognitron—. *The Transactions of the Institute of Electronics and Communication Engineers of Japan. A*, **62**(10):658–665, 1979.
- [13] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**(4):193–202, 1980.
- [14] Greenberg JH. A quantitative approach to the morphological typology of language. *International Journal of American Linguistics*, pages 178–194, 1960.
- [15] Hawrylycz M, Bernard A, Lau C, Sunkin SM, Chakravarty MM, Lein ES, Jones AR, and Ng L. Areal and laminar differentiation in the mouse neocortex using large scale gene expression data. *Methods*, **50**(2):113–121, 2010.
- [16] Hyvärinen A and Hoyer PO. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, **41**(18):2413–2423, 2001.
- [17] Kell A, Yamins DLK, Norman-Haignere S, Seibert D, Hong H, DiCarlo JJ, and McDermott JH. Computational similarities between visual and auditory cortex studied with convolutional neural networks, fMRI, and electrophysiology. *Journal of Vision*, **15**(12):1093–1093, 2015.
- [18] Ng L, Lau C, Sunkin SM, Bernard A, Chakravarty MM, Lein ES, Jones AR, and Hawrylycz M. Surface-based mapping of gene expression and probabilistic expression maps in the mouse cortex. *Methods*, **50**(2):55–62, 2010.
- [19] Oizumi M, Albantakis L, and Tononi G. From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. *PLoS Computational Biology*, **10**(5):e1003588, 2014.
- [20] Rothschild G, Nelken I, and Mizrahi A. Functional organization and population dynamics in the mouse primary auditory cortex. *Nature Neuroscience*, **13**(3):353–360, 2010.
- [21] Shinomoto S, Kim H, Shimokawa T, Matsuno N, Funahashi S, Shima K, Fujita I, Tamura H, Doi T, Kawano K, et al. Relating neuronal firing patterns to functional differentiation of cerebral cortex. *PLoS Computational Biology*, **5**(7):e1000433, 2009.
- [22] Terashima H and Okada M. The topographic unsupervised learning of natural sounds in the auditory cortex. In *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pages 2321–2329, 2012.
- [23] Terashima H. *Computational Model for Auditory Cortex: An Analogy to Visual Cortex*. PhD thesis, The University of Tokyo, 2014.
- [24] Tononi G. Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin*, **215**(3):216–242, 2008.
- [25] Whaley LJ. *Introduction to typology: the unity and diversity of language*. Sage, 1997.
- [26] Yamins DLK, Hong H, Cadieu CF, Solomon EA, Seibert D, and DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, **111**(23):8619–8624, 2014.