

W-3-4

ワークショップ「言語と非言語の時間生成—言語はなにをしているのか」

発表 4

言語の脳内情報表現と時間スケール

西本伸志

(大阪大学大学院生命機能研究科/NICT)

1. はじめに

ヒトは言語を用いることで対話相手との情報伝達、直近の状況認識、また現状に至るまでの文脈や背景の理解などを行っている。これら異なる時間スケールを持つ言語情報は、ヒト脳においてどのように表現されているのだろうか。発表者らのグループでは、動画を視聴する、対話をするなどの様々な体験をしている時のヒト脳を主な対象にして、体験内容と脳活動の関係性を調べることで、脳がどのように多様な情報を扱っているかを定量的に解き明かす研究を進めている。特に近年では機能的 MRI (fMRI) を代表とする大規模脳活動計測、また GPT などの大規模言語モデルを利用した言語情報の定量的な解析手法の発展により、言語に関する様々な情報を司る脳の情報表現やその機能構造についての研究が進展している。本発表では、これら最近の進展を支える基本的な考え方や方法論、および様々な時間スケールにおける言語情報と脳表現の固有の結びつきなどに関連した最近の研究等について紹介する。

2. 脳の情報表現と言語

私たちはその日常生活について視聴覚を代表とする感覚器から膨大な情報を受け取っており、それらの情報を適切に処理し解釈することによって周辺の状態を把握し、行動を行っている。その情報処理の物理的な実体は脳における神経活動（以後、脳神経活動）であり、様々な体験内容（風景を見る、話を聞く、文脈を推定する、等）と脳神経活動の関係性を調べることで、言語を含む私たちの多様な営みがどのような情報処理の結果として成立しているかを理解することが出来ると考えられる。

2.1 脳神経活動と体験内容

近年の脳計測技術の進展により、ドラマを視聴する、対話をする、などの日常的な体験を行っている際のヒトの脳神経活動を安全かつ高い解像度で計測できるようになった。例として機能的 MRI (functional magnetic resonance imaging; fMRI) といった装置を使えば、非侵襲的に（つまり、手術等を必要としない形で）ヒト全脳の脳神経活動を数 mm 角程度の解像度（成人の場合は約 5~8 万点）で 3 次元的に記録することが出来る（図 1）。たとえば研究者自身や学生被験者を対象にして映画 1 本分（数時間）程度の動画を見ている時の脳活動を fMRI で記録することで、様々な視聴覚体験に対応して時々刻々と変化する脳神経活動を記録することが出来る。

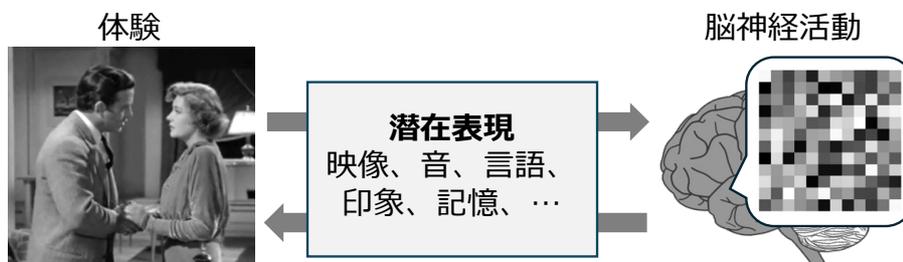


図 1 研究手法の概念図。様々な体験を行っている時の脳神経活動を対象に、体験内容と脳神経活動の関係性を司る潜在表現を明らかにすることで、脳が映像や音、また言語や記憶などの知覚や認知に関わる多様な情報をどのように表現しているか等について定量的な知見を得る。

上記のようにして得られた脳神経活動は、しかしただ目で見るだけで解釈することは難しい。数万点からサンプルされた脳神経活動が時間的に上下する様子を見ても、一見するとランダムにふるまっているようにしか見えない。私たちは、このような脳神経活動について、ヒトの体験内容が何らかの形で反映された、脳の暗号あるいは脳の言葉のようなものだと考える。このような立場からは、神経科学者の役割は、体験内容と脳活動の関係を司る暗号解読、あるいは翻訳を行うようなものと捉えることが出来る。より数理的な表現では、体験という時空間パターン（感覚入力、運動出力等）と脳神経活動という時空間パターンの関係を司る定量的な予測モデルを構築することができれば、これらの脳の言葉の理解に一歩近づけたことになる（Naselaris et al., 2011）。

2.2 言語特徴と脳神経活動

体験と脳神経活動の関係を理解するには、両者の関係性を司る何らかの内在的あるいは潜在的な特徴を捉えることが重要である。特徴とは、音声言語的な体験を例にすると、相手の話す聴覚入力に関する音の周波数、音素、単語、品詞、文構造、などが挙げられる。これらの言語の特徴としてヒトが知覚し扱うことの出来る既存概念については、これらに対応する何らかの神経活動が脳内に存在しているはずである。これらの特徴が脳のどこにどのような組み合わせや表現方法で処理されているかを調べることは神経言語学（neurolinguistics）の課題であり、視聴覚などのモダリティ普遍性や注意の影響なども含め、様々な研究が行われている（de Heer et al., 2017; Nakai et al., 2021）。

近年、いわゆる自然言語処理（Natural Language Processing; NLP）技術、特に 2018 年以降に発展した大規模言語モデル（Large Language Model; LLM）と呼ばれる言語モデルの研究と社会応用（ChatGPT 等）が進んでいる。大規模言語モデルは、コーパスの統計的な情報を階層的かつ数理的に扱うことで、従来のより小規模な言語モデルのように単語や単文の特徴を扱うだけでなく、複数文にわたるようなより長期的な関係性や、それらの意味内容等に関するより多様な言語特徴についても定量的な潜在特徴（ベクトル情報）として扱うことが出来る。これらの潜在特徴は脳神経活動とも対応があることが判ってきており、この性質を利用したより複雑な言語認知に関する研究が進められている。たとえば、脳内では現在までに得られた音声言語の内容だけではなく、未来に発せられるであろう発話内容の予測なども階層的に表現されている等の報告がなされている（Caucheteux et al., 2023）。

3 言語の脳内情報表現とモダリティ・時間スケール

上述のように、音声言語等を知覚している時のヒト脳活動を調べることで、言語を構成する様々な特徴が脳のどこでどのような形式で表現されているかが明らかにされつつある。一方で、ヒトが言語を用いて世界を理解する方式は、音声言語の利用に限らない。たとえばドラマや映画などを見ている際に、ヒトは演者が発している音声発話内容（たとえば「初歩的なことだよ、ワトソン君」というセリフ、図2）だけでなく、その際に視覚的に得られるシーンの客観的情報（2人の男が列車内にいて、座席に座りながら対話をしている。発話をした人物はタバコを吸っている）、そこにいたる背景や物語のコンテキスト（探偵シャーロック・ホームズがその友人ワトソンに対してこれまでの状況から考えられる推理を開示する、物語序盤のクライマックスである）など、様々な情報を言語の形式で扱うことが出来る。このような複数の感覚モダリティや異なる時間スケールの情報は、脳内においてどのように表現されているのだろうか。これらは同じ言語情報として脳内で一括処理されているのだろうか。あるいはそれぞれに対応する別々の脳領域や表現形式が存在するのだろうか。



図2 ホームズとワトソン
Sidney Paget illustration for
“The Adventure of Silver
Blaze”, Public Domain

3.1 動画視聴実験と多層的言語アノテーション

発話内容、状況の記述、物語上の展開といった異なる時間スケールの言語・意味情報の脳内表現を調べるため、発表者らのグループでは、様々なドラマや映画（Suits、Breaking Bad、Dream Girls など）を視聴している時の脳活動を fMRI で記録した。また、動画中の各シーンについて、発話内容、状況の客観的記述、物語上の展開などについての記述（言語アノテーション）を別途収集した。ドラマ・映画は被験者一人あたり 8 時間以上の視聴を行い、アノテーションは内容に応じて 1~5 秒間隔のサンプルとして収集した。またアノテーションを各種の大規模言語モデル（GPT、Llama 等）に入力し、その潜在情報表現について言語特徴の情報を表すベクトルとして捉え、脳活動との対応を多変量回帰（正規化線形回帰）によって求めた。

3.2 モダリティ・時間スケールによって異なる言語脳情報表現

上記の言語アノテーション等に由来する潜在情報ベクトルを刺激特徴として用いることで、ドラマ視聴下におけるヒト脳活動のモデル構築を行った。また、発話内容、客観的記述、物語上の展開に由来する特徴成分がそれぞれ固有に説明する成分について、Variance Partitioning 法を用いて定量してその脳内分布を求めた（図 3）。その結果、これらの 3 成分はそれぞれ固有の説明成分や脳内における担当部位を持つことが判った（Nakagi et al., 2024）。

これらの結果は、同じ言語情報でもその内容や時間スケールによって異なる情報として脳内で表現されていることを示唆する。また、単語単位の情報表現を示した言語モデル（Word2Vec）よりも文章単位の情報表現を表した大規模言語モデル（GPT、Llama など）の方がより優位な脳活動説明能力を示していることも判った（Nakagi et al., 2024）。これは、上記で用いた解析のフレームワークは、単語で表される即時的な情報だけでなく、より長期的な時間スケールの特徴を包含していることを示唆する。

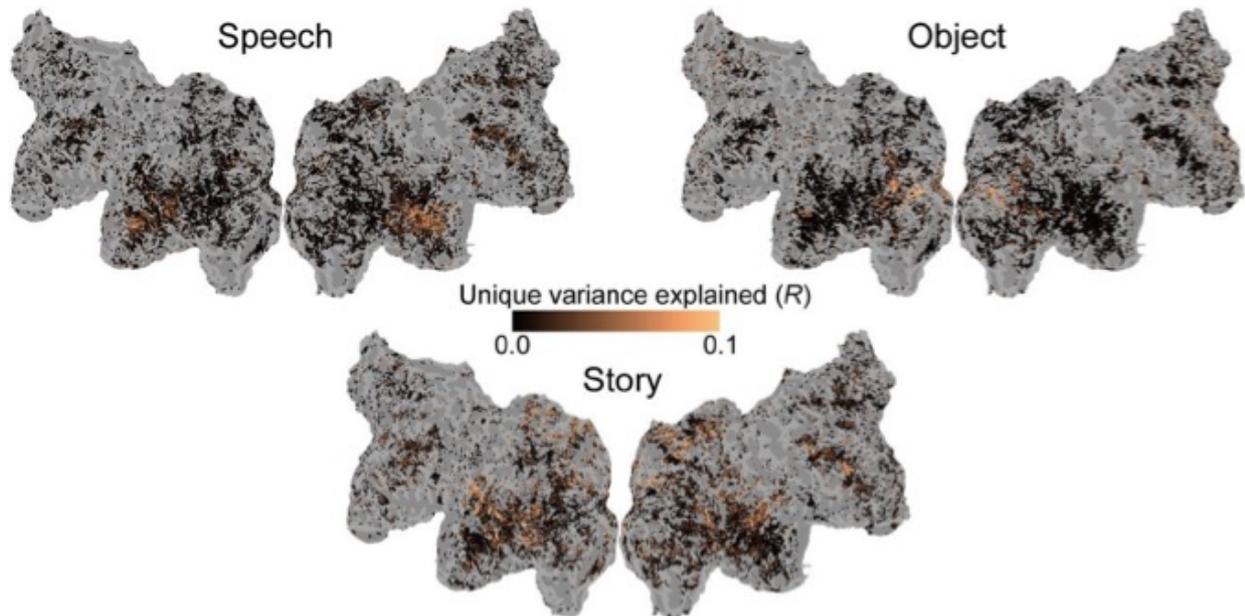


図 3 言語情報の脳内表現分布。動画を視聴している時の脳活動について、発話内容（Speech）、各シーンの客観的記述（Object）、物語上の展開（Story）といった言語で記述が可能な特徴成分について、それぞれが固有の説明能力を示した場所を色で表す。図は左右大脳を平面化したもので、中心部が後頭部、左右周辺部が前頭部を表す。発話内容は主に聴覚野が対応しているのに対して、物語は楔前部や前頭部などの広域で表現されている（Nakagi et al., 2024 より抜粋。CC BY4.0）。

4 おわりに

本稿では、言語を対象にした脳神経科学の方法論の例、および実例として発表者らが進めている最近の研究例の一部について紹介した。ここで紹介を行った神経言語学 (neurolinguistics) と呼ばれる分野は、比較的新しい分野でありながら、日進月歩の勢いで進展している。言語は脳の産物であり、また脳の仕組みの理解には言語は強力なツールになりうる。今後も言語学と神経科学が結びつくことで、言語とは何か、脳はどのようにそれを理解し生成するのか、時間や自己といった概念はどのように表現されるのか、といった多様な知覚・認知に関する理解の促進に期待したい。

引用文献

- Caucheteux, C., Gramfort, A., & King, J. R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, 7(3), 430–441.
- de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. *The Journal of Neuroscience : the official journal of the Society for Neuroscience*, 37(27), 6539–6557.
- Nakagi, Y., Matsuyama, T., Koide-Majima N., Yamaguchi H., Kubo, R., Nishimoto, S., Takagi, Y. (2024) The Brain Tells a Story: Unveiling Distinct Representations of Semantic Content in Speech, Objects, and Stories in the Human Brain with Large Language Models. *bioRxiv*, doi: 10.1101/2024.02.06.579077
- Nakai, T., Yamaguchi, H. Q., & Nishimoto, S. (2021). Convergence of Modality Invariance and Attention Selectivity in the Cortical Semantic Circuit. *Cerebral Cortex*, 31(10), 4825–4839.
- Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, 56(2), 400–410.