解答または解答例及び出題意図

年度	2026 年度
研究科	脳科学研究科
専攻・コース等	脳科学専攻
試験科目	外国語（英語）

～出題意図～

本問題は，Kumar et al. (2024)「Shared Functional Specialization in Transformer-Based Language Models and the Human Brain」(Nature Communications, 15, 5523) からの引用文を素材として，博士課程入学者として求められる英語学術読解力・概念的理解力・批判的思考力・研究者としての論述能力を総合的に評価することを目的としている。

1. 選文の背景と意図

本論文は，計算論的神経科学・自然言語処理・認知神経科学が交差する最前線の研究であり，Transformer の attention head という具体的な計算単位を用いて脳の言語処理を記述しようとする点で，方法論的・理論的な新規性を持つ。博士課程向けの試験素材として，高度な概念の組み合わせの理解力，脳科学研究科の研究内容との整合性などの観点から選定した。専門外の学生に対して不利益にならないよう，馴染みのない専門用語についてはその意味の解説を参考情報として問題文に記載してある。

2. 各問の出題意図

【問 1】和訳問題（下線 4 箇所）

4 箇所の下線部はそれぞれ，論文の論理的骨格を構成する命題であり，本文全体の主張の流れを代表している。語彙の置き換えではなく，文章の意味を自分の言葉として再構成できるかどうかを問う。

【問 2】概念的理解・論拠説明問題

本問は，表面的な内容要約ではなく，著者の主張の構造的論拠を再構成する能力を問う。「なぜ embeddings だけでは不十分か」という問いに対して，(i)理論的必然性（情報流という観点からの不可欠性），(ii)方法論的優位性（headwise 分解可能性），(iii)実証的証拠（脳活動予測における優位性）の三層で答えられるかを評価する。

【問 3】英語自由論述問題

博士課程の研究者として求められる能力を評価することを目的とした問題である。選択式とした理由は，受験者が自身の専門的素地（情報学寄り・神経科学寄り・言語学寄り等）に応じて最も深く論じられる問いを選べるよう設計したためである。

～解答例～

問 1

下線(a)

【解答例】

これらの知見は言語の神経生物学的基盤を築くうえで礎となったものの，実験室外への一般化可能性には限界があり，それらを自然言語の全複雑性に対応できる統合的モデルへと統合することは困難であることが判明している。

下線(b)

【解答例】

Transformer とは，再帰的な接続を用いず，多層の「attention head」回路を採用した深層ニューラルネットワークであり，大規模な実世界テキストコーパス上での自己教師あり学習を可能にする。

下線(c)

【解答例】

こうした取り組みは，もっぱら「embeddings」-Transformer による言語内容の表現-にのみ焦点を当てており，「transformations」-attention head が実行する実際の計算処理-をほぼ見落としてきた。

下線(d)

【解答例】

Transformer アーキテクチャは，過去の単語の意味が現在の単語の意味へとどのように取り込まれるかを定量化するための候補メカニズムへの明示的なアクセスを提供する。

問 2

【解答例】

著者が transformations を分析対象に加えるべきだと論じる根拠は，大きく三点に整理できる。

第一に，情報の流れという観点からの理論的重要性である。embeddings が各時点での意味の静的な「状態」を表すのに対し，transformations は単語間で情報がどのように流れ，現在の単語の意味がどのように更新されるかという「プロセス」を捉える。著者は，「どの単語間の統語的・文脈的情報が現在の単語の意味に影響を与えるかは，transformations を通じてのみ導入される」と明示している。すなわち，言語処理の動的な計算過程を記述するためには transformations が不可欠である。

第二に，機能的分解可能性である。embeddings は複数の attention head の変換が融合した結果であるため，どの計算がどの言語機能を担うかを分離できない。一方，transformations はヘッド単位（headwise）に分解可能であり，例えば動詞の直接目的語の解決や名詞修飾語の追跡といった，解釈可能な言語操作への機能的特化（functional specialization）が各ヘッドに確認されている。この分解可能性こそが，脳の言語処理との対応関係を精密に検討するための鍵となる。

第三に，実証的な優位性である。著者らの分析では，transformations は embeddings に匹敵する脳活動予測精度を示し，非文脈的 embeddings や古典的統語アノテーションを一般的に上回ることが示された。特に，モデルの早い層の transformations は embeddings そのものよりも脳活動の固有分散をより多く説明し，文脈情報がより早期の処理段階から豊富に存在していることが示唆されている。

問 3

選択肢(1)：Transformations と人間の言語処理の共通点・本質的差異

【解答例】

There are genuine points of overlap between headwise transformations in Transformers and human language processing. Both involve context-sensitive operations. Meaning is not fixed in isolation. It is shaped by surrounding linguistic context. The attention mechanism in Transformers captures this dependency explicitly. Similarly, human language comprehension involves predictive and integrative processes. The brain continuously updates word meaning based on prior context. Empirically, the paper shows that headwise

transformations predict brain activity in language regions. This suggests a functional alignment, at least at the representational level.

However, the differences are fundamental. Transformers process all positions in a sequence simultaneously. The human brain processes language in real time, in a strict left-to-right temporal order. This is not a minor implementation detail. It reflects a different computational logic entirely. Transformers have no intrinsic temporal dynamics or memory limitations. The human working memory system imposes hard constraints on integration distance. Moreover, human language processing is grounded in embodied experience, social interaction, and sensorimotor systems. Transformer representations are derived solely from distributional statistics over text. There is no referential grounding, no pragmatic intent, and no communicative goal.

A further difference lies in learning. The brain acquires language through developmental interaction with a social environment. Transformers are trained by predicting masked or next tokens in text. These are structurally different optimization pressures. Functional similarity in the final representational space does not imply mechanistic equivalence. The overlap observed in encoding models may reflect shared solutions to a shared computational problem. But it does not mean the underlying processes are the same.

選択肢(2)：Transformer とヒトの脳をより深く比較するために必要なアプローチ
【解答例】

Several advances are needed to deepen the comparison between Transformers and the human brain. Current encoding model approaches are a useful starting point. But they have important limitations that should be addressed directly.

First, the field needs causal rather than purely correlational methods. Showing that Transformer representations predict brain activity is informative. It does not show that the brain uses those representations. TMS and lesion studies can test whether disrupting a specific brain region selectively impairs computations analogous to particular attention heads. This kind of intervention evidence is currently missing.

Second, temporal resolution must be improved. Transformers process sequences in parallel. The brain processes language over time. EEG and MEG provide the millisecond-level temporal resolution needed to trace this unfolding. Combining mobile EEG with naturalistic speech stimuli would allow researchers to track which transformer layer best predicts brain responses at each moment in processing.

Third, the comparison should move beyond text to spoken and multimodal language. Current Transformer models are trained on written text. Human language is fundamentally spoken and embedded in social interaction. Speech Transformers and visually grounded models offer a more biologically plausible basis for comparison.

Finally, individual differences and developmental trajectories must be incorporated. The brain's language system varies across individuals and changes with age and experience. Comparing different Transformer architectures and training regimes against these differences could reveal which properties of the model are genuinely brain-like and which are artifacts of large-scale text training. This would move the field from analogy to mechanistic theory.