P-8

深層学習型 AI 言語生成が言語基盤研究に与える影響の評価

廉田（かどた） 浩（ひろし）

［要旨］深層学習型 AI 言語処理システム(AI 言語システム）により生成された文章が人間がみて自然で文法的にも適格である場合，言語基盤研究の諸課題に対して重大な影響を及ぼす可能性があることを示す。

最初に，AI 言語システムの基本構成とその要素技術を概観する。従来から提案されている脳内言語生成モデルでは，「語彙の記憶と使用」と「文法適格構文の生成」は別処理で，後者は文法専用論理処理機構で実行すると仮定しているのに対して，AI 言語システムでは，両者とも長時間学習でニューラルネット(NN) 内に形成された記憶の操作により実行している。次に，人間の脳内神経回路網と NN を比較し，少なくとも記憶形成に関して神経回路網は，より短時間で実現できることを説明する。これらの結果「刺激の貧困」状態でも人間脳では「文法習得」の可能性が想定できる。最後に，脳内言語生成モデルとして，従来の文法専用論理処理機構をもつモデルに代わる，AI 言語システムに近い記憶機構中心のモデルの提案を行う。

１．はじめに

深層学習型 AI 言語処理システム(AI 言語システム）が注目を集めている。(OpenAI, 2022 等）これらのシステムは，質問や要望を言葉で入力すると，そこからキーワードや文脈を読み取って，人間から見てもっともらしい回答をごく自然な文章の形で出力するので，社会的インパクトが大きいのが主な注目理由である。

一方，言語研究の分野では，主な研究対象が人間の生成する言語なので，応用面はともかく，基盤研究に対するこの種のシステム開発の影響は過小評価されるかもしれないが，本稿では，AI 言語システムの生成する文章の形式的な品質の高さを考えた場合に，脳内の言語生成モデルや「文法の生得性仮説」等の基本的課題(例えば，Jackendoff, 2002）に対しても重大な影響を及ぼす可能性があることを示す。

最初に，現時点における AI 言語システムの基本的構成と要素技術を概観し，これらの中核が画像認識等にも応用できる比較的汎用性の高い深層学習・記憶要素（ニューラルネット : NN）からなることを示す。また，従来から提案されている標準的な脳内言語生成モデルとの類似点や相違点を検証する。次に，AI 言語システム機能の飛躍的高度化の要因である NN と人間の脳内神経回路を形式的に比較し，その学習や記憶の特徴を対照する。最後に，AI 言語システムが生成する文章の形式的多様性や文法的適格性が十分検証されたとした場合，現在提唱されている脳内言語生成モデルに代わって，AI 言語システムに近いモデル，即ち，脳神経回路内に形成された記憶とのマッチング機構をベースした新モデルの提案を行う。

２．AI 言語システムの概略構成と(事前学習時を含む）入力情報の吟味

AI 言語システムによる文章生成の基本原理は「文章の途中に現れる単語は，直前・直後の複数の単語と，その文章の（キーワード等を含む）コンテキストから概ね予測可能」というものである。その予測機構の中核にある NN は，「入力情報に対しあるべき出力情報(正解情報）が得られるようにネットの結合係数を調整(事前学習）した大規模回路網」である。しかし，この基本原理を一括して実行できる単一 NN は超巨大なものになり，かつ非現実的なほど大量な事前学習を必要とするので実現困難である。そこで AI 言語システム

では，基本原理およびそれに対応する処理を二分割し，a)広域コンテキスト→局所コンテキスト予測，b)局所コンテキスト→単語予測，という二段階処理で文章生成を実現している。これらの概略を説明する。

2.1 AI 言語システム要素技術１：単語の分散表現（単語と局所コンテキスト処理）(Mokolov, 2013)

ある単語 X の局所コンテキスト $LC^{(X)}$とは，文中の X を中心として直前と直後に隣接する１～数語をとってきた語集合である。任意の単語 Xj に対して一般にコーパス中に M 個の使用箇所があり，対応して M 個の局所コンテキスト：$\{LC^{(Xj)}_i\}_{i=1\sim M}$ がある。中間段のノード数 d の NN に対して，入力集合を$\{Xj\}_{j=1\sim N}$ (N:コーパスの全異なり語数) 期待出力集合を$\{\{LC^{(Xj)}_i\}_j\}$ とした組み合わせ等を使って事前学習することにより，NN の中間段への結合係数 $\{w^j{}_k\}_{k=1\sim d}$ から，各単語 Xj に対する「分散表現」の d 次元ベクトル $\mathbf{v}^j$ を得ることができる。次元数 d はシステムによって異なり，通常は 100 以上である。この事前学習の所要時間は，高性能のコンピュータで数日以上かかる大規模なものである。また，単語情報を入力して分散表現を出力する操作は，入力と整合する記憶内容を出力する「記憶とのマッチング処理」と見なすこともできる。

ここで，単語の分散表現ベクトルの性質を少し考察する。分散表現ベクトルは，その生成方法から明らかに局所コンテキストの情報を含意している。従って，類似の分散表現ベクトルをもつ２つの単語は，類似の性質をもつ，一種の paradigmatic な関係にある。分散表現ベクトルの興味深い性質は，通常のベクトルと同様の演算や操作が可能な点である。(1) に加減算の例を示す。この例から推測すると，単語の意味や性質を規定する独立した因子は，分散表現ベクトルの別の次元に分離して記憶されていると考えられる。

(1) {King, Queen, Male, Female} 各分散表現ベクトルを {A, B, C, D} とすると，$\underline{A - C + D \fallingdotseq B}$

そして，このベクトルは，単語の語彙的因子だけでなく，文法的因子も含意していると考えられる。例えば，フランス語で，前方局所コンテキストに ‘la’ や ‘une’ を含む単語のベクトルでは，品詞と文法性の次元にそれぞれ「名詞」と「女性」のフラグが立っていると考えられ，また，日本語で，後方局所コンテキストに「接続助詞テ」を含む単語のベクトルでは，活用形の次元に「連用形」のフラグが立っていると考えられる。つまり，単語の分散表現には実質的に文法規則の一部(形態論的規則等) が記憶されているので，語彙の選択と同じ手続きで，文法規則検証用論理処理などを経ないでも，自動的に適格な文を生成できる可能性がある。

次に，２つの単語の因子別類似度を評価することを考える。前述の通り，類似の性質をもつ単語は類似の分散表現ベクトルをもつので，単語 Xa と Xb の各分散表現ベクトル(d 次元)を，$A=[a_1\ a_2\ ....a_d]$, $B=[b_1\ b_2\ ...b_d]$ とすると，Xa と Xb が類似の場合，(a_i , b_i) が共に 0 または 共に 正(または 非 0) のことが多いが，Xa と Xb が類似でない場合は，(a_i , b_i) が共に 正(または 非 0) のことが少なくなる。この結果，ベクトル A, B の内積　$A*B = \Sigma_i a_i b_i$ を計算すると，類似の場合は内積が正の値，類似でない場合は 0 に近い値になる。更にある因子の類似性に関しては，分散表現ベクトルのその因子に対応する次元部分だけを取り出して，部分ベクトルを作り(マスキング) それらの内積を計算することで，その因子の類似度を評価することができる。

また，ベクトルの加減算の代わりに，ベクトルを適切に回転させることによっても，別の単語の分散表現ベクトルを得ることが可能である。例えば「青(名詞)」の分散表現ベクトルの品詞部分を，名詞 → 形容詞 へ回転すると「青い(形容詞)」の分散表現に近いベクトルを得ることができると推測される。

このようなベクトルの演算や操作の組み合わせを使って，単語(の分散表現) 間の関係や文の構造を調べる Attention 方式や構文の Attention パターンなどについて次節で記述する。

2.2 AI 言語システム要素技術２：Attention と Transformer (広域コンテキスト処理) (Vaswani, 2017)

広域コンテキストとは，従来から使われているコンテキスト：文脈に近いもので，注目している文で使わ

れている全ての単語や，それ以前の文中の単語，および，各単語にともなう知識などに複雑に関係して形成される「状態」である。広域コンテキストを表現するものとして，過去には時系列情報処理システムの内部状態(Hidden-State) 等が想定されていたが，現在では「どの単語とどの単語がどのように関係しているか」を明示的に示す Attention 方式が多くの AI 言語システムで使われている。この方式で，Attention のパターン(厳密には，有向グラフ) には，実効的に文法規則(統語論的規則等) も含まれる。Attention パターンの例を図 1 ((Alammar, 2018)の例を多少変更したもの) に示す。

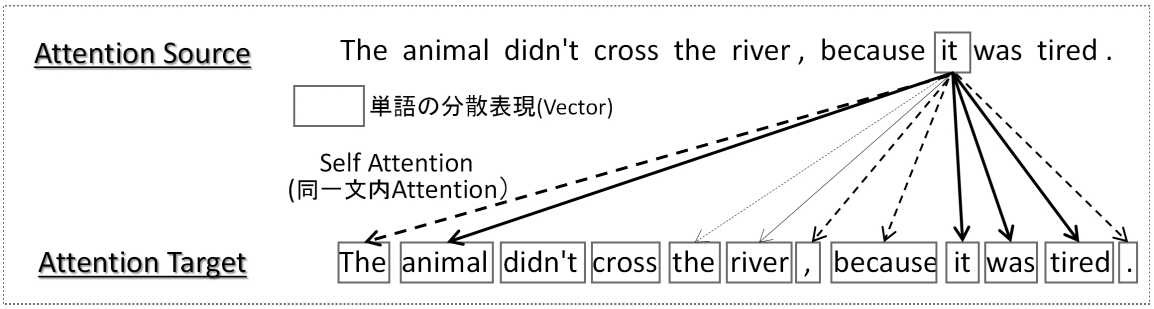

図 1．(Self)Attention パターンの例

広域コンテキスト(≒Attention パターン) を事前学習して文生成に使うためには，そのパターン種類数が計算可能な程度に制限されている必要がある。そこで，Attention パターンの学習を単語間ではなく，その分散表現間で行うことによりパターン種類数を大幅削減する。そして，このパターンの効率的な学習用に Transformer という機構が開発されている。Transformer ユニットの内部には Attention 情報学習用の NN が複数含まれていて，事前学習の結果がそれらの NN の内部結合係数に記憶される。AI 言語システムには，通常この Transformer ユニットが数十個配置され，それらを使った並列処理により学習処理が更に高速化される。Attention パターン学習の実行時間は，分散表現の学習時間と同程度（同オーダー）とされている。

次に，AI 言語システムを使った翻訳作業で，分散表現と Attention が協働し，意味と統語とが融合している文の解読・生成にある程度成功している例を示す。

2.3 AI 言語システムによる機械翻訳文の評価

英語文を日本語文に機械翻訳する作業で，英語文中の「代名詞 it の前方照応」が正しく解読されているかを確認する。但し，英語文は (Alammar, 2018)の例文を少し変えたもので，機械翻訳は “google 翻訳(Devlin, 2019)” を使用した。例文(2)~(4)の英語文の形式はほとんど同じであるが， because 節の述語部が異なっている。そのため it の照応先 (即ち，it の Attention 先) が，(2)では 主語の The animal であるのに対して，(3)では，目的語の the river になり，(4)では，前方照応が存在しない。一方これらの英語複文に対応する標準的日本語文は，従属(副詞)節と主節の主語が同一の場合，文頭に共通主語が来て，次に従属節述語が埋め込まれ，文末に主節述語という形式になる。主語が異なる場合は，文頭に従属節の記述があり，次に主節主語，文末に主節述語という形式になる。(2)~(4)の日本語訳文では，ほぼこの標準的規則に従った形式が採用されているので，少なくとも，(2)における前方照応は正しく解読されている。(3)では，「it≠animal」は解読済だが，「it=river」の解読は不明である。(4)の「前方照応不在」は解読済と推測される。

(2) The animal didn't cross the river, because it was tired.
⇒ 動物は疲れていたので川を渡りませんでした。

(3) The animal didn't cross the river, because it was a torrent (/ rapid stream).
⇒ それは急流だったので，動物は川を渡りませんでした。(ただし回答は安定していない)

(4) The animal didn't cross the river, because it was stormy.

⇒ 嵐だったので，動物は川を渡りませんでした。

分散表現と Attention による「代名詞の照合」解読のプロセスは，以下の手順で進められると推測される。

(5) a. it の局所コンテキストにより，その分散表現ベクトルを取得し，it の照応の有無を確定する

b. 前方照応ありの場合，学習済の Attention パターンによって照応先候補を選出する

c. 照応先候補のうち it の分散表現ベクトルに近い分散表現ベクトルをもつ候補を選択する

(2)の場合，it の分散表現ベクトルでは，照応有で，「名詞相当，有生性（LC= be tired が感情・感覚表現⇒主語は有生)」なので，この２つの特徴がそろった分散表現ベクトルをもつ The animal が照応先になる。

３．従来の人間の脳内言語生成モデル概要と疑問点

前節では，AI 言語システムにおける言語生成手順の概略を見てきた。その結果，これらのシステムでは，① 語彙記憶の形成・活用 と ② 統語や形態的変換といった文法処理 がともに，コンテキスト等の単語間の関係性の大規模な学習・記憶とのマッチング処理 によって実現されていることが分かった。

この節と次節では，人間の脳内での言語生成に関して，従来から提唱されている標準的な生成モデルを概観し，AI 言語システムと対比した場合に生じるいくつかの疑問点を述べる。

3.1 言語生成の下位機能分解と各専用処理ブロック

言語生成機構は,①語彙の記憶機構 ②統語形態等の文法処理機構 ③音声処理機構 という三種の下位機構に分解され，それぞれに専用処理ブロックが存在する，というのが従来の平均的な言語生成システムのモデルである。このシステムに対して，他の記憶（言語化以前のエピソード記憶）等から再生された情報が入力され，標準的な出力として，音声言語が得られる。図 2 にそのブロック図を示す。但し，各ブロックの厳密な役割分担や相互インターフェースというよりは，ブロックの概念を示したものである。

ここでは， ① と ② について検討する。特徴的なことは，語彙の記憶機構と 文法処理機構とが分離していることであり，特に議論があるのが「文法処理機構は生得的かどうか」という点である。

次に，この点に関しての従来の議論を整理してみる。

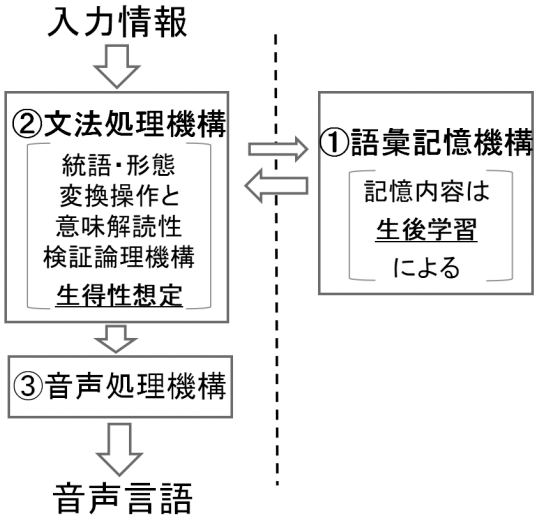

図 2. 従来の(脳内)言語生成モデル・ブロック図

3.2 文法規則の「学習」と語彙の「記憶」

脳内の言語生成で, 文法処理を実行している部分（ブロック）の構成は生得的であるという主張（文法機構生得説）は，以下のような根拠によるものと考えられる。

(a) 文法的に適格な文を生成するためには，文法規則に則った手続きで統語形態変換操作を行う必要がある。

(b) その文法規則を知るためには，非常に多くの文・文章を経験し「学習」しなければならない。

(c) 人間が生まれてから母語を獲得する（相当数の文法規則をマスターし，各種文法的適格文をしゃべれるようになる）までに，そのように多くの言語経験をするのは困難 ＝ 「刺激の貧困」

(d) この「刺激の貧困」状態を回避するためには，文法規則体系学習済の半固定的な処理機構が生得的(遺伝的）に備わっていることが必要。(ただし処理機能の小変更はパラメータ設定によって可能)

(e) 各国語の文法規則には，共通（普遍的）な部分と異なる部分があり，共通部分は上記生得的な処理機構をそのまま実行させ，異なる部分は，処理機構の機能変更用パラメータの設定（生後学習）で対応。

(f) 文法規則体系に従った半固定処理機構をもつため，規則の例外等に対応困難。(体系内不整合拒否)

文法機構生得説の主張でキーになる仮定は，(b) (c) (d) と考えられる。(e) の仮定は，文法機構生得説の一種の弱点である。(パラメータによって変えられる部分が増えると，文法規則はほぼ全て生後学習可能)

一方，脳内言語生成の語彙の記憶に関しては，生得的という主張は見られない。それは以下の事情による。

(g) 語彙は一項目ごとに「記憶」するもので，文法規則のように体系全体を「学習」するものではないので，多くの言語経験を必要としない。また，必要最低限の語彙量はそれほど多くなく，時間をかけて蓄積可能

(h) 語彙は，各国語で違いが大きので，もしも語彙記憶の生得性を主張すると，却って論理的に破綻する。

つまり，従来の生成モデルでは「文法処理のための事前『学習』は膨大(長時間）で，語彙の収集のための事前『記憶』は相対的に短時間で可能」である。一方，２節で述べたように，AI 言語システムでは，語彙記憶に相当する分散表現の事前学習，および，文法処理に相当する Attention パターンの事前学習は，いずれも膨大なものになっている。この学習時間の対比，そして，仮定(b) (c) (d) の必然性について次節で述べる。

４．人間の脳内の言語生成機構の推定

4.1 AI システムのニューラルネット（NN）と人間脳内神経回路網の各記憶特性の対比

AI 言語システムにおける NN は，元々人間の脳内神経回路網(通称「ニューロナルネット」，本稿では「神経回路網」と略称する）を簡略化して模したものであり，同じような動作特性をもつ場合と，異なる動作特性を持つ場合とがある。例えば，あるパターン集合の事前学習をする場合，２節でも説明したように，NN では学習結果はネット内部の結合係数変化という形で記憶される。一方，神経回路網に対する学習結果は，部分的に結合係数変化に記憶されるが，多くの記憶は結合係数そのものではなく，むしろそれにより新たに形成される内蔵ループにダイナミックな形で記憶されると考えられる。ここでは大量のループを内蔵することを強調するため「L 神経回路網」と呼ぶことにする。ループ内蔵回路網の特性は非線形回路理論で説明可能だが本稿では省略する。この違いのため，記憶形成に関して，両方式で以下のような違いが発生する。

(x)記憶形成に要する時間（＝事前学習時間）は，NN に比べ L 神経回路網では大幅に短縮される。

(y)新たに記憶項目を追加する時，NN では以前の記憶内容を全面的に書き換えなければ，記憶項目の一部を喪失する場合があるが，L 神経回路網では追加書き込みでもそのような部分喪失がほとんど発生しない。

次に，我々が実際に経験する各種記憶の特性について考える。多くの記憶の特性は，上記の L 神経回路網の特性に近いものである(Busáki,2022)。例えばエピソードの記憶では，一回のエピソード経験でその記憶が形成される。語彙記憶も同様に，一回の辞書参照等で形成可能である。また，一旦形成された記憶に対して更に一項目の記憶を追加しても，それによって以前の記憶の一部が失われるということがほとんどない。(即ち，記憶の追加蓄積が容易）本稿では，このタイプを「パターン記憶型」と呼ぶ。一方，別のタイプの記憶も存在する。代表的なものが，手続き記憶（例えば，自転車に乗ること）であり，記憶形成に時間がかかる。特性面だけで分類すると，このタイプは NN に近く，本稿では「手続き記憶型」と呼ぶ。

4.2　人間脳内言語生成モデルの文法処理機構生得性の再考と新モデルの提案

図 2 に示した従来の言語生成モデルでは，語彙記憶機構はパターン記憶型の処理であり，文法処理機構は手続き記憶型の処理を専用論理処理機構化したものと考えられる。

一方，AI 言語システムの分散表現生成部は，語彙記憶機構に近い機能をもつと考えられるので，この分散表現生成部での NN を L 神経回路網に置き換えれば，(x)の記述のとおり事前学習の時間が短縮され，(g)のような特性をもつことになるので，脳内言語生成モデルの語彙記憶機構の具体的な構成と考える事ができる。

次に，文法処理機構であるが，まず「文法処理」自体を再考してみる。(a)の記述通り，文法処理は文法的に適格な文の生成のための操作であるが，実は，このために明示的な文法規則を知ることが必須ではない。むしろ，文法規則を学習するために使ったコーパス内の多種多様な例文の形式リストを全て記憶できれば，そのリスト中から照合・選択して文生成した方が，より確実に適格文が得られる可能性がある。そして従来は，この全形式リストが膨大なものになり，実現困難と考えられていたが，AI 言語システムでは，単語の分散表現という「Paradigmatic な単語の小集合」単位で各例文形式の記述を行い，情報の種類数を大幅に削減してリストを作ることで，明示的に文法規則を知ることなく適格文の生成を可能にしている。これと同様のこと，即ち，従来の語彙記憶の代わりに単語の分散表現が記憶され，従来の文法処理の代わりに Attention パターンが記憶されているシステムが人間の脳内にあり，しかも両方の学習記憶用として，NN ではなく，L 神経回路網が使われているとすれば，膨大な事前学習を必要としないで，文法的に適格な文を生成するシステムができあがることになる。つまり，(b)や(c)の状況は発生しないので，生後の学習で各記憶は獲得でき，文法処理機構が生得的である必要性はなくなる。これに伴い，(f)の状況も発生しないので，多少の文法体系の不備(文法規則間の不整合や例外等) は許容される。このことから，文法体系全体知識は必須ではなくなる。

4.3　新モデルにおける通言語的文法共通性の解釈

従来の通言語的文法共通性は，「人間は遺伝的に同じ文法処理専用機構を持って生まれてくるので，各国語の文法体系の基本部分は共通のもの（＝普遍文法）になる」という説明がされている。「文法規則に従ったアルゴリズミックな文生成が実行される」のではなく，本稿で提案したような「各種構文形式記憶の照合と選択」による文法処理機構を仮定した場合，通言語的文法共通性（の少なくとも一部）は，文法専用処理機構というより，その中核にある，人間の記憶システムの共通性に起因すると考えられる。ただし現時点で，具体的に「記憶システムのどの特性がどの文法規則に対応するか」と特定するレベルには達していない。また，人間と人間以外の動物とでは，事象の記憶の仕方が異なっていることが知られているので(例えば，Quiroga, 2023)，これが「人間言語とそれ以外の動物『言語』との違いの原因」との推測も可能である。

５．まとめと今後の課題

本稿では，最近注目を集めている，深層学習型 AI 言語生成システム(AI 言語システム)が生成する「文」が十分な多様性を持ち，かつ文法的にも適格であると仮定した場合，「人間脳内言語生成モデルや言語生成機構の生得性有無」といった言語の基盤研究にどのような影響があるのかを文献調査などにより概観した。更に，脳内の言語生成モデルとして，従来の文法規則に準拠したアルゴリズミックな論理処理ではなく，AI 言語システムの方式に近い記憶ベースのものに変えたものを提案し，その特性を推察した。従来の脳内言語生成モデル，AI 言語システム，新提案の言語生成モデルの概略比較を表 1 に示す。

表１．各種言語生成モデルの生成方式・特徴の比較

内部構成	項目	従来の脳内モデル	AI言語システム	新提案脳内モデル
語彙記憶	方式	脳内神経回路網による記憶のアクセス（詳細アクセス機構不定）	単語の分散表現（局所コンテキスト）記憶のアクセス*)	単語の分散表現〝と類似の〟記憶のアクセス*)
	方式示差特徴	知識の追加容認	知識の追加容認	知識の追加容認
	記憶媒体	脳内神経回路網	NNの結合係数	脳内L神経回路網
	事前学習	中（「刺激貧困」なし）	長大（巨大知識集積目標）	中（「刺激貧困」なし）
文法処理	方式	文法専用論理回路でアルゴリズミックに生成（詳細内部構成不定）	Attentionパターン（広域コンテキスト）記憶とのマッチング**)	Attentionパターン〝と類似の〟記憶とのマッチング**)
	方式示差特徴	文法体系内不整合拒否	文法体系不備容認	文法体系不備容認
	記憶計算媒体	脳内神経回路網	NNの結合係数	脳内L神経回路網
	事前学習	生得（「刺激貧困」回避）	長大（文法体系網羅目標）	中（「刺激貧困」なし）

アクセス*) ⇒ 一種の連想アクセス　　マッチング**) ⇒ アプリにより具体的操作が異なる

本稿で提案した脳内言語生成モデルが有効であるためには，AI 言語システムが文法的に適格な全ての構文形式を生成できることが必要であり，現時点では，このことは検証されていない。今後の第一の課題として，AI 言語システムにより生成される文や文章を多面的に検証することは，言語学の基盤研究にとってだけでなく，AI 言語システムの実社会応用面からも極めて重要であるので，その推進が期待される。

また，新提案の脳内言語生成モデルの有効性がある程度確認された場合，言語の変化，特に文法体系の変化はどのような要因で発生するのかを検討する必要がある。基本的に新提案モデルの文法処理は，既存の文や文章の集合の生後学習から得られた情報に従って行われるので，文法体系変化は頻繁に起こるとは考えられない。但し，従来モデルと異なり，新提案モデルでは文法体系の不備もある程度許容されるので，文法体系 AA から BB へ変化する場合，その変化が時間をかけて進行し，変化途中で，AA → AB → BB といった，AA 体系と BB 体系の交雑体系 AB（体系内整合性がとれない場合がある）が長期間現れたり，極端な場合，交雑体系 AB で変化が終了する場合も許容される。

参考文献

Jackendoff, R. (2002), *FOUNDATION OF LANGUAGE*, Oxford Univ.Press, 郡司隆雄(訳) *言語の基盤* 2006 年 岩波

Mikolov, T., et al.(2013) “Efficient estimation of word representations in vector space” *ICL Workshop 2013*

Vaswani, A., et al.(2017) “Attention Is All You Need” arXiv:1706.03762 (https://arxiv.org/archive/cs.CL)

Alammar, J. (2018) “The Illustrated Transformer” Blog, https://jalammar.github.io/illustrated-transformer/

Devlin, J., et al. (2019) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” arXiv : 1810.04805v2 [cs.CL] (Computer science > Computer and Language)

OpenAI(2022), “Introducing ChatGPT” https://openai.com/blog/chatgpt

Buszáki, G., et al. (2022) “Neurophysiology of Remembering” *Annu. Rev. Psychol.* 2022.73:187-215

Quiroga, R. Q., (2023) “An integrative view of human hippocampal function: Differences with other species and capacity considerations,” *Hippocampus Vol.33 Iss.5* May 2023 p.616-634