

令和 6 年 6 月 9 日現在

機関番号：13801
研究種目：挑戦的研究（開拓）
研究期間：2021～2023
課題番号：21K18115
研究課題名（和文）脳科学・認知科学による人間に近いモデルに基づく日本語話し言葉解析器の構築と検証

研究課題名（英文）Japanese language processing system by human-like model based on brain and cognitive science

研究代表者
狩野 芳伸（Kano, Yoshinobu）
静岡大学・情報学部・准教授

研究者番号：20506729
交付決定額（研究期間全体）：（直接経費） 19,900,000円

研究成果の概要（和文）：本研究の目標は、脳科学の知見を導入したより人間に近い言語処理モデルを構築することである。理論言語学の知見に基づきfMRI実験の課題をデザインし、構文や語順を題材に実験を行った。広くAI分野で用いられているTransformerとその亜種では内部で個別情報をどう処理しているか定かでないため、言語学的な機能との対応を分析した。Transformerには明示的なメモリ（記憶）がなく、必要とする学習データ量も人間に比べ桁違いに大きく冗長な可能性があること、またその内部表現の解釈が困難なことから、リカレントネットワークの記憶に明示的な解釈を強制したモデルを構築し実験を行った。

研究成果の学術的意義や社会的意義
人間の仕組みからは遠い人工的なモデルが工学的応用の主流となっている現状に対し、脳科学・認知科学と理論言語学、自然言語処理の接点を見だし3者の知見を融合した新たなモデルを構築する糸口とすることができた。このような試みでは、より大規模な工学的モデルと個別タスクでの性能比較によってその時々では劣って見えてしまうことが多いが、「より人間に近い言語モデル」という理学的モチベーションは、長期的なブレークスルーの可能性を見据えた「人間との親和性」「説明可能AI」という工学的側面においても意義があり、広く社会的意義のある研究成果といえる。

研究成果の概要（英文）：The goal of this study is to construct a language processing model that is more human-like by incorporating insights from neuroscience. We designed fMRI experimental tasks based on theoretical linguistics insights and conducted experiments focusing on syntax and word order. Since it is unclear how individual information is processed internally in Transformers and their variants, which are widely used in the AI field, we analyzed their correspondence to linguistic functions. Transformers lack explicit memory and require an exponentially larger amount of training data compared to humans, potentially leading to redundancy, and their internal representations are difficult to interpret. Therefore, we constructed and experimented with a model that forces explicit interpretation of memory in recurrent networks.

研究分野：自然言語処理

キーワード：自然言語処理 脳科学 認知科学

1. 研究開始当初の背景

自然言語は人間の脳機能に基づいており、表層的な情報を用いた大規模データを前提とする現在のアプローチの限界を超えるには、言語処理過程やその制約をも、より人間に近づけることが必要であると考えられる。現在の機械学習手法の多くは任意の関数を近似しうる表現能力をもっているが、たとえば人間の構文処理ではありえない構造も表現可能であり、特に生成においてその影響が如実に表れ得る。このことは構文に限らず意味など他の属性でも同じで、人間の脳構造に由来する制約に近づけなければ、言語で意思疎通する際に必須である曖昧性の適切な解消も難しいはずである。もうひとつ端的に問題が現れるのが話し言葉である。話し言葉は省略が多く、誤認識や誤発話もあり、発話の途中でも内容の理解が必要であるうえ、自然言語処理の基本単位である「文」の区切りすら曖昧である。人間の脳神経回路網はそのまま実装できるレベルには解明されていないが、現在のアプローチではこのように既知の言語脳科学、言語心理学的知見ともかけ離れた、人間とは根本的に異なる処理モデルを用いている状況であった。

研究代表者は、代表者である科研費萌芽研究「人間同様に失敗する構文処理による自然な文生成の研究」「心理学的に妥当な自然言語処理システムによる会話の自然な「間」の自動生成」において、人間に近い構文処理と文生成の研究を音声および話し言葉を対象に進めてきたが、既知の実験的知見のみに頼ったモデルの構築であった。

ここには、人間の逐次的処理、時間軸での処理過程の動作と記憶の利用や制限といった脳科学からの知見が直接的に結合しうる。本研究は、自然言語処理・言語脳科学・理論言語学の研究において、個々の専門家を結集させるという特色があった。

2. 研究の目的

本研究では、自然言語処理と脳科学の研究を融合させより人間に近い言語処理モデルを構築すると共に、言語解析器を実装する。言語学・科学(人間の言語処理モデルの解明)、工学(解析器の実装)が繋がった新たな研究を創成するのが目的である。

3. 研究の方法

言語野には「文法・読解・音韻・単語」から成る4つの言語中枢が同定されている。また、分担者(酒井)らの研究で文法関連ネットワークは少なくとも3つあることが明らかとなった。機能的MRI(fMRI)などによる脳機能計測と機能結合(functional connectivity)の解析により、これら3つの脳内ネットワークがどのように融合されているかを明らかにする。fMRI実験の課題は、分担者(福井)を中心に、理論言語学の知見に基づきデザインする。脳の言語情報処理では、書き言葉の視覚入力話し言葉の聴覚入力と明確に分離しているが、文法処理と意味処理においては両者が完全に重なると予想されるため、入力情報を統合した言語中枢の動作原理を明らかにできる。分担者(酒井・福井)の共同研究により文法中枢の動作原理が既に示されているので、残る意味処理の部分の部分を明らかにする。

自然言語処理分野においては、生成AIの急速な性能向上をはじめ、Transformerをベースとした大規模言語モデルが、分野内の様々なタスクにおいて利用されている。Transformerが内部に持つSelf-Attention機構(注意機構)は、自然言語処理以外にも様々な分野において効果的な手法として利用されているが、それぞれのAttentionモジュールに対しての解釈は不明な点が多い。そこでTransformerモデルの内部動作を分析するため、Attention機構をヘッド単位で言語学的な機能と対応させることにより、Transformerモデルの内部の挙動を解析する、脳科学の知見を活用した手法を用いる。具体的には、特定の品詞にあたるトークンを入れ替えたり、語彙を固定して構文的な順序のみを変更する変換を行い、変換前後でのAttention Headの反応の差分を観察する。

さらに、人間らしい言語モデルをつくるため、人間らしい構文処理の特徴を2つ仮定したモデルを構築する。1つ目は逐次処理である。人間は聴覚視覚から受け取った情報を受け取ったものから順に処理している。1つの文を理解するとき、文末までの情報を一度すべて並列処理するのではなく逐次処理するのが適切であると考えられる。この点に対応するためRNN構造を持つLSTMをベースに新しいモデルを作る。2つ目は構文解析候補数を削減することである。特に話し言葉では、さまざまな省略や挿入が発生するが、人間のように適切に推測して補完するには候補数1の確度の高い推測が必要である。袋小路文に対して人間が混乱するのは、途中まで入力された情報からすでに構文解析候補を1つ持っていて、次の文字列を入力されたときに持っていた候補が誤りであることに気づくからであると考えられる。候補数が1になるという点は、記号化の導入によって再現を試みる。

具体的には、モデルの計算を人間の処理に近づけることと説明可能性の向上を目指し、逐次処理を行う LSTM における記憶セルで記号表現を扱う構造として、LSTM-Symbolic を提案する (図 1)。分散表現と記号表現の変換は Embedding 層の重みをそのまま利用する。記憶理想ベクトルを導入し、逆伝播時の計算を独自に定める。また、2 層の LSTM 言語モデルと LSTM 内での記憶の記号化によって、記憶の制限付き文脈自由文法を仮定したモデルを作成する。

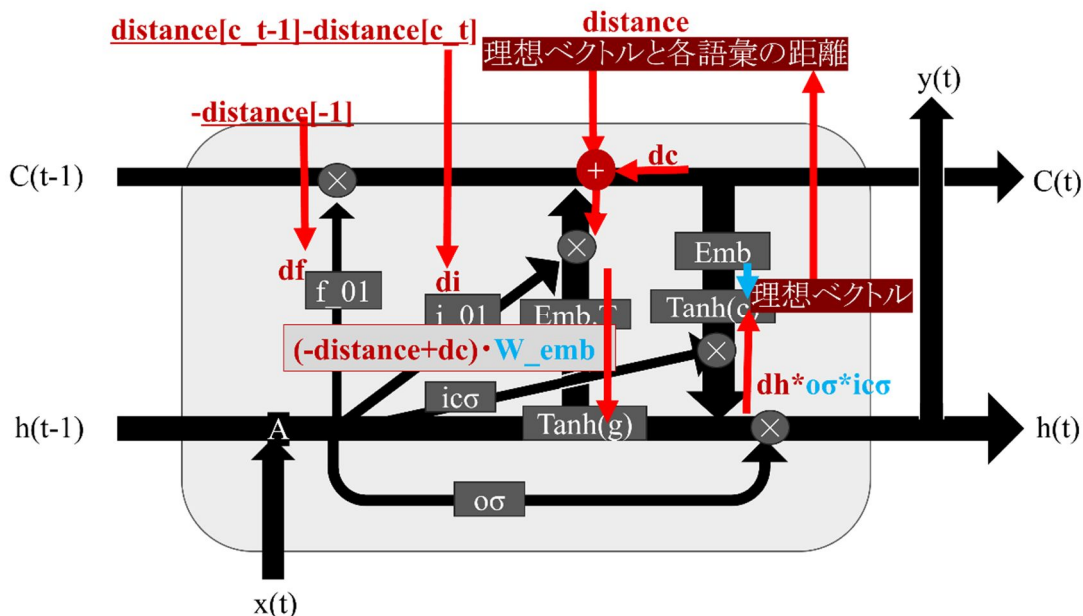


図 1. LSTM モデルの記憶部分に記号化の制約を加えた提案モデル

4. 研究成果

fMRI を用いた実験に関しては、自然言語処理技術を用いた実験データ作成を行い、構文的な機能の差分を確認する被験者実験を行い、その結果を下記の工学的実験にフィードバックした。

Transformer の内部動作分析については、以下のいくつかのパターンで実験を行った。まず、ある文の特定の品詞 (接続詞および格助詞を対象とした) にあたる非内容語を、同じ品詞の別の非内容語で置換した。原文と置き換えられた文との Attention ノルムの差分をとることで、置き換えられた非内容語に共起する Attention Head を取得した。文ペアにおける Attention の差分を各レイヤー、各ヘッドについて取得し、この差分を箱ひげ図によって可視化して、特定の品詞の非機能語における Attention の分布の違いにより特定の品詞の非機能語にかかわる機能に対応する Attention Head の特定を試みた。

さらに、隣接しない文節に係り受けがあった場合、係り受けを置き換えた文を作成した。係り受けを変えた際に文意が不適切になる文章に対しては MLM Scoring を用いて、統計的にありそうな文章かどうかを判定し、文意が通ると思われる文章のみを残すフィルタリング作業を行った。なお実験には 12 層からなり 1 層あたり 12 個のヘッドを持つ BERT モデルを使用した。以下では入力に近い側から第 1 層、第 2 層... と呼ぶ。

結果、格助詞・接続詞・係り受け関係それぞれに対して、レイヤー 7・8・9 層目において差分の出方には違いがあるものの、顕著な差が見られた。これらの品詞や文章構造においては共通し

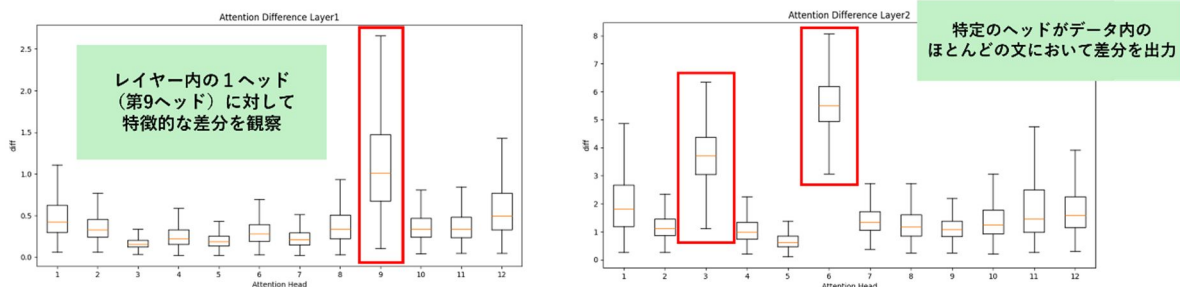


図 2. 格助詞「が」を別の格助詞に置換したときの第 1 層の Attention ノルムの差分の分布 (左)
係り受け構造を変えた文ペアの第 2 層の Attention ノルムの差分の分布 (右)

た Attention Head が作用し、何かしらの特徴を捉えている可能性があり、相当な言語学的役割分担があるのではないかと考えられる[1]。図 2 に分析結果の一例を示す。

より人間に近い言語モデルの実験では、LSTM-Symbolic モデルを次単語予測により学習して生成された文章の構文の妥当性で評価し、記憶の記号化による効果を調べた。その結果、同じ層数でも入力層側に記号化の構造を持つモデルは有意な差でほかのモデルの評価を下回った。また、提案モデルは文頭の扱いという局所的に望ましい生成例が見られた。提案モデルの方が重みのパラメータ数が少ないにもかかわらず、ベースラインと同等の性能を達成したことは、より構文処理に本質的な構造を捉えられたことを示唆している[2]。

将来展望として、より大規模なデータで実験を試み、性能向上を図ることが挙げられる。また、モデルが何層、何次元であるべきか、記号化する語彙はどのようなセットであるべきかといったことをさらに探求したい。

1. 馬場 海好, 狩野 芳伸. **日本語事前学習済み BERT の言語学的機能による差分入力を用いた Attention ヘッド別機能分析**. 2024 年度人工知能学会全国大会 (JSAI2024). アクトシティ 浜松. 2024/5/30
2. 小川 麻美, 狩野 芳伸. **記号と分散表現の統合のための記憶に制約を課した RNN の次単語予測による事前学習**. 2024 年度人工知能学会全国大会 (JSAI2024). アクトシティ 浜松. 2024/5/30

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 上山 彩夏, 狩野 芳伸	4. 巻 37(2)
2. 論文標題 単語統計を損失関数に取り入れた深層学習による多様な雑談対話生成	5. 発行年 2022年
3. 雑誌名 人工知能学会論文誌	6. 最初と最後の頁 G-L62_1-10
掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.37-2_G-L62	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 Sakai, K. L., Oshiba, Y., Horisawa, R., Miyamae, T., Hayano, R	4. 巻 32
2. 論文標題 Music-experience-related and musical-error-dependent activations in the brain	5. 発行年 2022年
3. 雑誌名 Cereb. Cortex	6. 最初と最後の頁 4229-4242
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/cercor/bhab478	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 Naya, Y., Sakai, K. L	4. 巻 16
2. 論文標題 Editorial: Task-related brain systems revealed by human imaging experiments	5. 発行年 2022年
3. 雑誌名 Front. Behav. Neurosci.	6. 最初と最後の頁 889486, 1-2
掲載論文のDOI (デジタルオブジェクト識別子) 10.3389/fnbeh.2022.889486	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -
1. 著者名 福井直樹	4. 巻 161
2. 論文標題 日英語比較統辞論とパラメータの理論(会長就任講演)	5. 発行年 2022年
3. 雑誌名 言語研究	6. 最初と最後の頁 1-33
掲載論文のDOI (デジタルオブジェクト識別子) 10.11435/gengo.161.0_1	査読の有無 有
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 -

1. 著者名 福井直樹	4. 巻 161
2. 論文標題 日英語比較統辞論とパラメータの理論 - FukuiとKurodaのシステムの相違点と類似点	5. 発行年 2022年
3. 雑誌名 言語研究	6. 最初と最後の頁 1-33
掲載論文のDOI (デジタルオブジェクト識別子) 10.11435/gengo.161.0_1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Sakai Kuniyoshi L, Oshiba Yoshiaki, Horisawa Reiya, Miyamae Takeaki, Hayano Ryugo	4. 巻 -
2. 論文標題 Music-Experience-Related and Musical-Error-Dependent Activations in the Brain	5. 発行年 2021年
3. 雑誌名 Cerebral Cortex	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1093/cercor/bhab478	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計2件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 室町俊貴, 狩野芳伸
2. 発表標題 生成モデルによる傾聴応答タイミングの推定と動的Prompt-Tuneを用いた応答スタイルのパラメーター制御
3. 学会等名 言語処理学会第29回年次大会 (NLP2023)
4. 発表年 2023年

1. 発表者名 上山彩夏, 狩野芳伸
2. 発表標題 ゼロ照応解析に基づく項省略補完を取り入れた対話応答生成
3. 学会等名 言語処理学会第29回年次大会 (NLP2023)
4. 発表年 2023年

〔図書〕 計5件

1. 著者名 Hiroki Narita, Naoki Fukui	4. 発行年 2022年
2. 出版社 Routledge	5. 総ページ数 344
3. 書名 Symmetrizing Syntax	

1. 著者名 酒井邦嘉（編著），合原一幸，辻子美保子，鶴岡慶雅，羽生善治，福井直樹（著）	4. 発行年 2022年
2. 出版社 中央公論新社	5. 総ページ数 217
3. 書名 脳とAI - 言語と思考へのアプローチ	

1. 著者名 酒井邦嘉（監修），日本科学協会（編）	4. 発行年 2022年
2. 出版社 中央公論新社	5. 総ページ数 328
3. 書名 科学と芸術 - 自然と人間の調和	

1. 著者名 遊佐典昭、小泉政利、野村忠央、増富和宏 編	4. 発行年 2023年
2. 出版社 開拓社	5. 総ページ数 400
3. 書名 言語理論・言語獲得理論から見たキータームと名著解題	

1. 著者名 酒井 邦嘉、合原 一幸、辻子 美保子、鶴岡 慶雅、羽生 善治、福井 直樹	4. 発行年 2022年
2. 出版社 中央公論新社	5. 総ページ数 224
3. 書名 脳とAI	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	酒井 邦嘉 (Sakai Kuniyoshi) (10251216)	東京大学・大学院総合文化研究科・教授 (12601)	
研究分担者	福井 直樹 (Fukui Naoki) (60208931)	上智大学・言語科学研究科・教授 (32621)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------