

dwango

DWAL-TR-2015-002

Jun. 9 2015

# 第9回 全脳アーキテクチャー勉強会 「実世界に接地する言語と記号」報告書

土屋敏明\*1      礼王懐成\*2      森田和彦\*3      山川宏\*4  
Toshiaki Tsuchiya      Kaisei Reio      Kazuhiko Morita      Hiroshi Yamakawa

\*1 株式会社 DHC シアター  
DHC Theater Co., Ltd.

\*2 メタップス  
Metaps Inc. Data mining Director

\*3 株式会社 ネットプロテクションズ  
Net Protections, Inc. Marketing Director

\*4 株式会社 ドワンゴ ドワンゴ人工知能研究所  
DWANGO Co., Ltd. Dwango Artificial Intelligence Laboratory

 ドワンゴ  
人工知能研究所  
DWANGO ARTIFICIAL INTELLIGENCE LABORATORY

# 第9回全脳アーキテクチャ勉強会

## 「実世界に接地する言語と記号」報告書

### The 9<sup>th</sup> Whole Brain Architecture workshop report

土屋敏明<sup>\*1</sup>  
Toshiaki Tsuchiya

礼王懐成<sup>\*2</sup>  
Kaisei Reio

森田和彦<sup>\*3</sup>  
Kazuhiko Morita

山川宏<sup>\*4</sup>  
Hiroshi Yamakawa

<sup>\*1</sup> 株式会社 DHC シアター  
DHC Theater Co., Ltd.

<sup>\*2</sup> メタップス  
Metaps Inc. Data mining Director

<sup>\*3</sup> 株式会社 ネットプロテクションズ  
Net Protections, Inc. Marketing Director

<sup>\*4</sup> 株式会社 ドワンゴ ドワンゴ人工知能研究所  
DWANGO Co., Ltd. Dwango Artificial Intelligence Laboratory

全脳アーキテクチャ勉強会では、毎回脳の機能の一部に注目し、神経科学、機械学習等の関連分野専門家の講演により、研究の現状や方向性について知識を共有し、脳機能の実現方法について整理を行っている<sup>1</sup>。これまでの人工知能や機械学習における現実世界のデータと記号や言語とのつながりは、人間に比べて極めて貧弱なものであった。しかし、近年においては、神経科学の知見が急速に蓄積され、深層学習においては基本的な概念の獲得に成功するだけでなく、文法を扱う試みなども始まっている<sup>2</sup>。言語を扱う知能を実現するためには、高次の概念やその関係を抽出し、記号と結びつけた上で、文法等に従った記号操作を行う能力が求められると思われる。しかし、こうした能力を工学的に実現するのは容易ではないと考えられる<sup>3</sup>。そこで第9回目の勉強会では「実世界に接地する言語と記号」をテーマとして、林隆介氏(産総研)、谷口忠大氏(立命館大)、酒井邦嘉氏(東大)をお招きし、2015年2月4日(水)夜にグラントウキョウサウスタワーにて開催した勉強会の様子を報告する。

## 1. 勉強会概要

本勉強会では、山川宏氏(ドワンゴ人工知能研究所)のオープニングに引き続き、林隆介氏(産業技術総合研究所システム脳科学研究グループ)から「脳内視覚情報処理における物体表現の理解を目指して: Deep neural network の利用とブレイン・マシン・インタフェースへの応用」と谷口忠大氏(立命館大学情報理工学部)による「記号創発ロボティクス: 内部視点から見る記号系組織化への構成論的アプローチ」のご講演を頂いた。その後の休憩をはさんで酒井邦嘉氏(東京大学大学院総合文化研究科)からは「脳科学から見た言語の計算原理」のご講演をいただき、その後全体討論をおこなった。

## 2. オープニング(山川宏氏)

### 2.1 講演内容

全脳アーキテクチャ掲げるミッションは、脳全体の部品を機械学習モジュールに置き換えることにより、世界初の汎用人工知能を構築することである。今年「全脳アーキテクチャ・イニシヤチブ」という NPO 法人を立ち上げることで、関連人材の育成などを軸に広く公益的な立場で全脳アーキテクチャの研究アプローチの普及を進める<sup>5</sup>。

現在、ドワンゴと理化学研究所の共同研究では、複数の脳の器官に対するマルチモジュールを統合するためのプラットフォームを開発している。その狙いは、プラットフォームを介して多くの研究者が開発した機械学習モジュールとして集結することであり、少なくとも以下の3つの研究プレイヤーが必要である。

- ① 脳全体を統合する回路を設計する認知アーキテクト
- ② 個々の器官に対する機械学習モジュールの開発者、
- ③ プラットフォームを中心としたソフト基盤の設計、改良を行う研究開発者、

脳器官に対応する複数の機械学習装置を合するにはいくつかの研究課題が想定されている。まず、個々の器官内で機能や現象の階層性やネットワーク等に対する取り扱いの課題がある。例えば、海馬という全体のモジュールに対して、海馬内の細かい機能をサブモジュールとして扱っていく必要がある。その際に、モジュールやサブモジュールの階層のスケールを変えながら開発していく必要がある。また全脳アーキテクチャでは、同じ器官やその内部機能に対してモデルが複数の場合は、任意に切り換え可能なように、各種接続条件の制限の元で開発していく必要がある。さらに、部分と全体を統合しながら開発や実験ができるプラットフォームや、脳の組織間の接続モデルを記述する専用言語とか、それを効率よく実行するプログラム等を含む開発環境も必要となる。

こうした研究者間での共通プラットフォームを構築するアプローチには、主に下記のような批判がある。しかしながら研究連携により開発を加速しうる本アプローチは有望な選択肢であると考える。

- 既存プラットフォームで十分であり、再開発の必要があるのか？
- 全脳を機械学習としてモジュール化し、さらに接続できるのか？
- 開発者が神経科学知見についての理解が不十分な段階で開発に着手するのは時期尚早ではないか？

ドワンゴ人工知能研究所は、上記の他に、新皮質-海馬連携モデル、新皮質が持つ汎用的な情報表現の取り扱い等について研究開発を進めている。その成果の一部は、直近では6月の函館で開催される人工知能学会の全国大会で発表する。

なお現在は AI もしくはニューロコンピューティングは第3次ブームとされ、知識表現の獲得についてある程度の目処が立つ

たといわれている。このブームにより進歩した技術が、全脳アーキテクチャを推進する技術的サポートになっている。なお今後も、ビッグデータ分野ではデータ量や多様性が減る事はなく、深層学習の分野では、応用面としてライブラリの利用技術が進み、基礎的な部分では時系列やトップダウン処理の方向に進展するであろう。つまり全脳アーキテクチャの実現可能性を高める技術の進展は続く見通しと考えている。一方、ハードウェアにおいては、現在の半導体技術には微細化の限界があるため、2020年頃には演算速度や記憶容量に頭打ちすると予測されている。そのため、単なる高速化・大容量化ではなく、知能技術への期待が高まっているという側面もある。

人工知能技術を進歩させつつ、技術的特異点となる人間に近い知能に到達するには、少なくとも真の「言語理解」という大きな壁を乗り越える必要があるため、その手前で、人工知能は再度「冬の時代」を迎える可能性もある。しかし最終難関ともいべき壁を越えた人工知能が現れれば、多くの SF でも描かれているように、その人工知能は世界中のあらゆる文章化された知識を取り込めるようになる。こうした背景から今回は、「実世界に接地する言語と記号」というテーマで勉強会を企画した。

### 3. 脳内視覚情報処理における物体表現の理解を目指して:Deep neural network の利用とブレイン・マシン・インタフェースへの応用(林隆介氏)

#### 3.1 概要(事前公開版)

視覚情報は、大脳皮質の各視覚領野において階層的に処理され、物体認識実現すると考えられている<sup>6</sup>。しかし、中～高次視覚野でどのような情報符号化は未解明である。本発表では、コンピュータビジョンの性能を飛躍的に向上させた、深層学習の一つである大規模な多層ニューラルネットワーク(DNN, Deep Neural Net)を用いたアプローチ<sup>7,8</sup>について述べる。DNN 各層の視覚情報表現とサルの大脳から記録した神経活動パターンを比較することで、DNN の高次層の情報表現が、下側頭葉の神経細胞による情報表現に類似していることを示し、さらに、自然言語における物体概念表現の類似性を示した<sup>9</sup>。こうした結果をもとに、視覚的メタファーの神経基盤について議論するとともに、神経情報から想起している視覚イメージを復号化するブレイン・マシン・インタフェース(BMI, Brain Machine Interface)技術への応用研究を報告する。

#### 3.2 講演内容

##### (1) 研究目的

研究目的は、脳の神経科学の基礎的な解明とその応用である。基礎的研究として、網膜が受け取った視覚情報を脳神経細胞が処理する仕組みを計算論的に解明する事である。現在、視覚情報についてコンピュータ上のディープニューラルネットワーク(DNN)と、動物モデルの大脳から記録した神経活動とを比較することで、脳内視覚情報表現の解析を行っている。脳における視覚情報処理を応用して、記録した神経活動から視覚情報を解読して制御するBMI(Brain Machine Interface)の開発に繋がるものと考えている。

##### (2) 実験:設定と結果

実ニューロンと人工のニューラルネットワークとの対応について実験した。

- 人工物としての一般画像認識用 DNN (Deep Neural Network)

ハードウェアとして、市販の GPU を1枚搭載した Linux PC を用意し、ILSVR2012 でトロント大から発表された Deep Learning の一種で、Convolutional Neural Network 用のライブラリである Cuda-convnet を利用して、5 層のコンボリューション型演算層と 3 層の全結合型演算層で構成される DNN を実装した。学習データは 120 万枚、テストデータは 15 万枚を使用し、1000 種類の分類課題でエラー率 16%の判別精度を達成した ILSVR2012 のベストモデルとほぼ同性能の DNN を用意した。

##### • 実ニューロンの脳神経回路網

比較的大きな脳を持ち、行動課題も学習可能な、人と同等の視覚機能を有するマカクザルを動物モデルとして利用した。脳の新皮質に、一辺が 4mm、電極数 96 本を持つマイクロ電極アレイを多数埋め込んだ。視覚物体情報は腹側視覚経路に沿って、並列階層的に処理されるため、下側頭葉の前部、中間部、後部にアレイ電極を埋め込んだ(合計電極数:224 本)。電極を埋め込んだ覚醒行動下の状態で、顔、身体部位、動物、植物、人工物などのカテゴリを含む 120 種類の画像を提示し、神経活動記録を実施した。

##### • 実ニューロン計測

計測結果は、120 種類の顔画像に対する応答が、一つの電極から得られる。100 本の電極から記録される神経活動を、100 次元ベクトルの時系列データとして、120 種類の各画像を高次元ベクトル空間の一点として表現して計測データを整理した。この結果は、単純な画素レベルの類似性ではなく、物体の意味情報を表現している。

##### • DNN の各層と実ニューロンの活動パターンの比較

実ニューロンの活動パターンを機械学習により特徴抽出し、DNN 各層における人工ニューロンの活動パターンとの相関を解析した。DNN 高次層の特徴を利用したときに、推定精度の向上が顕著であることがわかった<sup>10</sup>。

- 神経活動から DNN 高次層への写像を利用した画像復元  
提示画像に対する実ニューロン活動パターンを DNN の情報表現に投射し(ニューラルネットワーク7層、4096 次元の画像表現)、DNN が大規模画像データベース(120 万枚)から抽出、復元する画像の復元精度を分析した。その結果、実ニューロン活動から推定した画像と DNN の参照画像の類似度を比較では、Wordnet(英語の概念辞書)に基づく意味類似度評価から、高次層の視覚特徴を利用することで、意味的に類似した画像が抽出されることがわかった。

##### (3) 実験における課題

今後の展望として、実際のニューロンの応答を教師信号としてディープネットを学習させることを考えている。現在よりも大規模な画像セットを見せた場合の神経活動パターンを、そのままディープネットでも学習させて、最終的に生じる情報処理を理解する事が、本質的に重要である。そのために、脳に埋め込む電極数を4、5個に増やす予定である。記録画像は、今回の発表では、120枚のデータセットだった。今後は動画像や、イメージネット(10000 枚以上)、他には動画を自由観察する際の神経活動記録を行っていく予定である<sup>11</sup>。

##### (4) 視覚情報と、概念事象

今後は、視覚情報と単語のような概念事象の関連についても解明を進めたい。

関連研究では、人に動画を見せて fMRI データをとる実験では、動画の内容を予め細かくラベリングすることで、個々の脳領域における視覚的物体カテゴリに対する応答をマッピングでき

る。これらの神経活動パターンを利用することで、ラベルした単語に対する脳活動表現がカラーで表される。脳の活動パターンであるこれらの色を、例えば Wordnet<sup>12</sup>のネットワーク越しに見た概念の構造にラベリングすることで、ネットワークと色が非常に良く対応していることが分かってきている。つまり、辞書で見たときの概念と脳パターンが非常に良く類似していることが明らかになっている。

こうして概念と神経表現の対応が見えつつあるので、この対応付けの仕組みや規則について掘り下げていくことが今後の一つの大きな方向と考えられる。具体的には、実ニューロンの応答パターンを解析すると、通常視覚的に類似した画像同士や、物体カテゴリとして類似した画像の応答パターンの間で高い相関を示す。一方、手と果物とか、視覚的に似ても似つかないものに反応するニューロンが出てきたりする。これは、視覚的類似性だけでは説明できない。

単語の意味としては無関係ながら、経験的に共起しやすいパターンがある。例えば、「芝生を走る犬」という典型イメージである。すると、犬、ボール、芝生といったものが、常に共通の記憶に構築され、も連づけられて想起されるため、これが神経ネットワーク中に反映されると推測している。

### (5) 視覚研究の方向性

視覚の神経科学研究では、言語に基づく概念との融合研究が、今後の重要なアプローチとなるだろう。あるシーンを視覚刺激として入力し、その神経活動を記録することで、シーンに含まれるオブジェクトの要素群を明らかにして、対応する言語や概念のデータベースを解析すると、イメージやシーンの説明のようなものが再構成できると考えられる。

関連して、Google のグループ<sup>13</sup>が画像認識ネットワークの出力と、言語生成のニューラルネット(リカレントニューラルネット)を繋げて、教師あり学習により、画像のキャプションを自動生成する技術を報告している。

### (6) メタファーとDNN

言語や概念について、一つのキーワードとしてメタファーがある。経験から概念を構築していく過程で、類似する概念表現に重ね合わせながら、新しい概念を構築し、積み上げていく、という仮説である。ディープネットには、メタファーで表現される概念も表現できる可能性があると考えている。例えば、視覚的メタファーの具体例として、ある種の「見立て」や「物まね」がある。DNN による視覚情報処理におけるネットワーク最上位層のニューロンの働きを見ると、あるニューロンは「女性」と「フラミンゴ」とか、視覚的に類似したものが反応することがあり、メタファーが成立する仕組みを考えるヒントとなる。ここから、経験を一つの概念体系にまとめていく過程で、類似したものは概念として統合していくという普遍的な性質があるのではないかとこの言語解析に関する仮説が考えられる。こうした解析が人工的なニューラルネットでも可能になっていくものと予想する。今後も概念やカテゴリと視覚情報との関連について研究していきたい。

### (7) BMI(Brain Machine Interface) と視覚情報復号技術

神経活動を記録して、ディープネットを利用して、概念と結びつける技術は広い応用範囲をもつと考えており、それが BMI 技術である。例えば、視覚情報を主として BMI 技術を応用したロボットハンド制御<sup>14</sup>や、3次元の視覚情報を復元して、ユーザーにフィードバックする技術、さらに将来的には、脳情報との通信というところまで繋げていきたいと考えている。DNN の高次層の場合、ソフトウェア上のニューロンの情報表現が、生体ニュー

ロン活動とよく対応するということが分かってきており、BMI の視覚情報復号技術に有用であり、現実問題に直結するといえる。

### (8) 視覚研究と概念の理解

神経科学全般の問題意識<sup>15</sup>としては、高次元空間で表現されるニューロンの情報処理は人間にはとらえきれないので、DNN を介して理解する以外の手段は無いとも思われる。しかし、それだけだとブラックボックスになってしまうので、中で何が起きているのか、数理的なツールが必要と考える。もう一つは、学習のあとで、特定の機能を明らかにできるようなデータセットを実験者のセンスで入れる事で、より詳細に実行されている情報処理を解明していくプロセスが必要になると考えられる。視覚の研究者から見ると、テキストを含めた概念の理解が今後の鍵になってくる。最終的には、情報の処理だけでなく、実際の脳の記録を通して、脳とコンピュータを繋ぐことが重要ではないかと考えている。

### 3.3 質疑応答

- Q. 講演中で「ネットワークに与える情報は生体ニューロンの情報を与えるのが本質ではないか？」と話されたと思うが、その理由は何か？
- A. 学習した DNN のパターンへ生体神経活動のニューロンパターンを写像するときに、今のところ、回帰でフィッティングしている。この場合、生体ニューロン同士の空間的結合を個別に考慮しない形で、フィッティングしていることになる。そのため、例えば、同じ TE 野に埋め込んでいても場所によって全然違う機能の表現があったとしても、その機能差がフィッティングのときに潰れてしまう。これは DNN が生体反応とは独立にネットワークを学習している限界ではないかと推測している。これを回避するために、初めから1つの生体ニューロンがどう応答するのかを、DNN で学習させ、そのうえで、このネットでは何が起きているか、近隣のニューロン同士の情報処理とかを詳細に解析できるのではないかと考えている。

## 4. 記号創発ロボティクス: 内部視点から見る記号系組織化への構成論的アプローチ(谷口忠大氏)

### 4.1 概要(事前公開版)

人を含んだ創発システムの中でも重要なものに記号創発システムがある<sup>16</sup>。記号系とは形成された概念や語彙、文法、語用といった言語的知識を含む。これらは、人間の身体的・社会的経験に基づいた認知発達過程を通して、また、記号論的な相互作用を通して、ボトムアップに形成される創発的存在である。このような記号創発システムのダイナミクスを構成論的に理解することは、人間とコミュニケーションするロボットを作る上でも、人間の発達過程や文化を理解する上でも重要である<sup>17</sup>。本発表では、ロボットを用いた記号創発システムの議論を行うために、我々が展開する記号創発ロボティクスの研究領域について概説する。また、具体的な研究テーマとしてのマルチモーダル物体概念形成及び言語獲得や、時系列データの二重分節解析に関する研究成果に関しても報告する<sup>18</sup>

### 4.2 講演内容

#### (1) 研究動機

人のコミュニケーションを支える知能を理解したい。人間の知能は、生まれた時は、右も左も分からない。「右」という概念と「左」という概念が分化していない、未分化な認識世界の中で行動を

始める。そして、環境適応的に概念や行動を獲得していく。その延長線上の中で、言語を用いたコミュニケーションを獲得していく。その構造、計算論的プロセスを知りたいと思う。

## (2) 記号創発ロボティクス

- 谷口家の長男と次男の例  
人間の子供は、教師あり学習のラベルデータを学習用に入力していないにも関わらず、自律的に学習し縫いぐるみのような柔軟物体を右手で持ち上げたり、何かしらのプランニングを行ってプラレールを創造的に組み替えたりしていく。2歳くらいになると、設計者が SSH(Secure Shell)でログインしてプログラミングしていないにも関わらず、喋り始める。この能力はロボットの世界ではあり得がたい事象といえる。
- 言語の獲得プロセス  
人間は視覚・聴覚・触覚情報などのセンサー情報から言語を獲得する。視覚だけで数百万次元ベクトル、皮膚データを含めても数百万から数千万次元ベクトルデータをリアルタイムで処理していく。その感覚刺激・運動刺激から自己組織化的なプロセスにより言語を獲得するというプロセスを人間は体現している。この実現機構に対する理解が強い研究モチベーションである。
- 概念獲得と意味  
言語獲得の理解では、概念獲得と意味の扱いが重要となる。1960年代から議論されてきた人工知能研究において、意味はどのように扱われてきたか。有名なのが「意味ネットワーク」である。オブジェクト指向のような表現により、人工知能に記号間の関係性として言葉の意味を教え込む。しかし、ある言葉を別の言葉で表現すると、最終的には端点かループに陥る。例えば、「自由」という言葉は、「社会秩序の中で束縛されずに、自らの意思に従って生きること」と辞書では述べられている。では、「意思」とは何か。辞書では「積極的な心の持ち方」と述べられている。では、「心」とは何か。辞書では「～な精神作用」と述べられている。では、「精神作用」とは何か。辞書には述べられていない。端点に辿り着いた例である。また、「石」という言葉は、「砂より大きく、岩より小さい塊」と述べられている。では、「砂」とは何か。「非常に粒の細かい石」と述べられる。では、「岩」とは何か。「大きな石」と述べられている。ループに陥った例である。最終的には、言葉の意味は「感覚的に」理解するしかない。(鈴木孝夫「ことばと文化」参照)
- 記号接地問題  
人間が設計した記号に対して、ロボットがセンサ・モータ系を通していかに意味づけるか。人間が持つ概念や言語は、感覚によって支えられている。しかし、ロボットと人間の感覚は異なる。歴史上、人間は自らと異なる身体系を持つシステムとコミュニケーションをした経験がない。外国人の場合でも、言葉は異なるが、視覚・聴覚・触覚などの感覚器官は共有している。すると、ロボットに人間と同じ記号を押し付けるとは妥当か、という問いも必要になってくる。
- 人間は客観的な世界など見ていない  
人間は、視覚・聴覚・触覚などの感覚情報を介した閉じた世界を見ている(ユクスキュルの「生物から見た世界」参照)。認知的に閉じた経験から得られる情報の中でいかに記号系が組織化されていくのか。自らの閉じた世界に立脚して、多様な行動や概念を獲得し、それに基づいてコミュニケーションを行う知能を計算論的に理解したい。
- 記号創発システム

創発とは、システム構成要素間の局所的な相互作用を通じ、「大域的な秩序」が「ボトムアップ的」に発現し、こうしてできる「大域的な秩序」が境界条件として「要素間の局所的相互作用」をトップダウン的に支配するという双方向性の過程により、新しい機能・形質・行動などの獲得をもたらすことを指す<sup>19</sup>。

大域的な秩序とは、集団が持っている記号系や、コミュニティ内の相互信念のことである。そして、集団内の秩序から逸脱したコミュニケーションは許されない。例えば、リンゴをドナドナと呼ぶことにしても、お店では意味が通じず、リンゴを手にするにはできない。このようなトップダウンな制約が存在する。

一方で、新しい言葉を作ることもできる。例えば、全脳アーキテクチャ。かつては存在しなかったが、一定数の集団から受け入れられれば、新しい言葉として成立する。このようなボトムアップな自己組織化も存在する。

以上のように、「トップダウンな制約」と「ボトムアップな自己組織化」の中で記号が形成されるシステムを、記号創発システムと名付ける。

### • 記号創発ロボティクス

ロボットを人間のコミュニケーション内に存在させるには、記号創発システムの要素をロボットに組み込むことに他ならない。このように、記号創発システムを前提として、構成論的アプローチにより、ロボットに認知・運動・言語を学習させる研究を記号創発ロボティクスと名付ける。

ファーストステップとして、教師なし学習により、物体の概念を覚えさせることを目標としよう。

## (3) マルチモーダルカテゴリ形成

### • マルチモーダル情報統合

人間は、手触りなど様々な情報を統合して概念を獲得している。ロボットにも、身体性に基づくボトムアップなマルチモーダルカテゴリ形成ができるのではないか(マルチモーダルカテゴリ形成とは、例えば視覚・聴覚・触覚情報を用いたカテゴリ形成を指す)。

### • 実験内容

ロボットにぬいぐるみを視覚・聴覚・触覚情報により理解させた。結果として、多くの被験者と同様のカテゴリ分類を自動的に生成した。単一モダリティ(例えば、視覚情報のみ)による分類の場合は、分類精度が落ち、推定カテゴリ数も少なくなった。また、トピック数が固定の MLDA よりもトピック数が変更できる MHDP のほうがよかった。

### • 階層マルチモーダル LDA を用いたロボットによる階層的な概念の形成

hMLDA を用いて、マルチモーダルデータから物体の階層カテゴリの自動生成を行う。hMLDA は、物体を文書とみなし、視覚(SIFT)・聴覚・触覚・単語情報のそれぞれの特徴量を単語ととらえて、hLDA のようにノードを階層化したトピックモデルを構築している。hMLDA はトピックが1つのパスによって表現されている木構造から構成されて、それぞれのトピック間における特徴量の共有の割合が多いほど上位のノード、つまり抽象度が高くという特徴により物体の階層的な分類へと応用することができる。

### • Bag of Multimodal Hierarchical Dirichlet Processes (BoMHDP)

物体の概念だけでなく、色、形、などのマルチモーダル情報のコンテキストに依存したクラスタリングを行うことが目的。ロボットに言葉で教唆を行い、その状況での行動やセンサ

一情報から物体の様々な性質を自動分類し学習する。視覚(SIFT, 色を独立の特徴量として扱う), 聴覚, 触覚などマルチモーダルな情報に対して, それぞれ異なった重みの値(0, 1, 2, 5, 10, 20, そして 30)を組み合わせたMHDP モデルを作成する。BoMHDP はそれらMHDP2400個により構成されている。BoMHDPの特徴量であるマルチモーダル情報の重みの組み合わせに基づいて分類されたカテゴリと言語情報(ロボットに指示する言葉)の相互情報量を計算する。その相互情報量が最も大きい単語を対応するカテゴリのラベルとする。モデル間の距離をもとにMDSを計算した3Dでマッピングした結果, 色, 触覚, 楽器そしておもちゃなどの異なる視点でのクラスに分類されたことが確認された。

#### (4) 二重分節解析器

- 二重分節解析器とは  
音声認識の研究分野において, 人間の音声言語は「二重分節構造」を持ち, 意味を担う最小単位(形態素)と, それを構成するための最小単位(音素)の2種類からなるとされる。この構造が, 人間の動作データや, 自動車運転挙動データにおいても, 二重分節構造が潜んでいると考え, この構造を教師なし学習で推定する二重分節解析の手法を提案してきた。
- 二重分節解析器の応用  
二重分節解析器の応用分野として, 身体動作データ, 自動車運転挙動データ, 音声発話データ(HDP-HMMは使用せず, NPYLMだけを使用)がある。ここで, HDP-HMMは, “Hierarchical Dirichlet Process Hidden Markov Model”である。NPYLMは, “Nested Pitman-Yor Language Model”であり, 文字列から, 隠れた単語分割を推定しつつ直接, 言語モデルを作成する方法である。二重分節解析器の応用により, 自動車の運転など複雑な操作の系列をチャンク化して, 言語として解析することで, 人間の運転としての教師無し抽出ができる。すると, 運転者の意図が変わるタイミングが分かる。つまり, 予測も出来るようになる。
- Double Articulation Analyzer (DAA)  
HDP-HMMで時系列情報から無意味な音素(アルファベットなど)を同定し, NPYLMでチャンクを推定して意味のある語彙を抽出する。この手法を身体動作の時系列に適用して, 人間やロボットの動作を解析する。
- sticky HDP-HMM  
HDP-HMMは可変個(潜在的に無限個)の隠れ状態を持つHMMであり, 隠れ状態がすべての状態に遷移できる(ergodic HMM)という性質を持つ。自己遷移確率を高くなるようにバイアスしたSticky Hierarchical Dirichlet Process-HMMをもちいることで, 振動的なノイズのような高頻度な状態の遷移がなくなり, 安定的な持続的な状態を保持することができる。
- 教師無し形態素解析としてのNPYLM  
HPYLMはデータ中の各単語が生成されたN-gram長を隠れ変数とみなしてベイズ推定を行うことで, 文脈により様々な可変長N-gramの生成モデルを提案する。NPYLMはそのHPYLMをもとに可変長の単語N-gramモデルや文字N-gramモデルをiterativeに推論することで, 教師無し形態素解析を実現している。
- 運転挙動データの二重分節解析  
同様にHDP-HMMとNPYLMを運転挙動に適用して, HDP-HMMは運転行動の単位をセグメント化し, NPYLM

はセグメント化した運動行動の単位をチャンク化して文脈情報を持った運動系列を生成する。これによりチャンクの切れ目が, 例えば右折から直進のように運転行動の文脈の変化点と対応していることがわかった。そのチャンクの切れ目の前後だけを抽出して運転動画の自動要約に応用する。

- 運転挙動データからのトピック抽出  
各運転シーン内での挙動特徴, 画像特徴, 運転シーントグの分布からmultimodal LDAを用いて運転トピックを推定することにより, ドライバーと自然に協力できる運転支援システム構築のための人が付与した運転シーントグと融合し, これを基に運転挙動と前方画像から運転シーントグを予測することができる。
- 運転行動変化タイミングの予測  
HDP-HSMM(Hidden semi-Markov model)を用いてチャンクの変化点を予測する。HDP-HSMMにより推定された隠れ運転ラベルの持続時間分布パラメータを用いてNPYLMでチャンクの持続時間分布を計算し, 観測データから次のチャンクの切り替わり時間を予測できるようにした。

#### (5) 教師なし語彙獲得

- 教師なし語彙獲得  
マルチモーダルなトピックモデルを用いて形成された物体概念と教師なし形態素解析を結合した言語獲得を行う。発話を音節認識器で解析し, NPYLMを用いて教師無しで音節列を単語に分割し, 切り出された単語とparticle filter based online multimodal LDA (PFoMLDA)でトピック解析をしたトピック概念とを結びつけることで, その語意の学習を行った。  
また, 音声認識には音節認識のみだけでなく, 言語モデルを扱うことで物体名を正しく認識する精度を上げることができる。
- 自己位置推定と語彙獲得の統合  
ロボットが音声で指示された言語情報と場所の特徴を解析し, 場所に関する語彙獲得を行い, それを自己位置推定に有効活用する。具体的には言語情報として, 空間的な広がり属性を場所とし, 特定の座標や局所的な地点を位置とするなど語彙を使い分けながら, 自己位置推定法の一つであるMonte Carlo Localization法と掛け合わせて, 自己位置推定と語彙獲得の統合を行う。

#### (6) まとめ

記号創発ロボティクスの研究では, 記号はボトムアップに形成される創発的な存在であり, こうした記号に基づいて人類はコミュニケーションを行うことを前提とする。その上で記号を伝える認知, 運動, 言語, 学習についての研究を進めている。

## 5. 脳科学から見た言語の計算原理(酒井邦嘉氏)

### 5.1 概要(事前公開版)

言語は人間の脳に備わる本能であり, 文法に基づいて文の「木構造」が生成される。これまでの脳研究で, 文を理解している時の脳活動を測ることにより, 文法に特化した場所(文法中枢)が明らかになっている<sup>20</sup>。最近我々のグループは, 文法中枢を含む複数の脳部位(言語野)の活動が, 文の木構造の「併合度」に従って変化することを見出した<sup>21</sup>。講演では, 脳における言語の計算原理について議論したい。

## 5.2 講演内容

### (1) 脳における言語計算

- 言語は計算である  
言語の処理は、入力→分析→文法→合成→出力という一連のフローになっていると考えられるが、文法はブラックボックスになっている。文法という分野を脳科学的に証明したいというところから研究が始まった。人間で因果関係を実証するため、脳に何らかの損傷のある患者を測定することで研究した。サルなどの動物は人間のような言語能力を持っていない<sup>22</sup>。チョムスキーによれば、言語の文法的な性質は文法を計算する計算機械と対応付けられる。コンピュータ(チューリング機械)が持つ計算能力をどこまで制限すれば、人間の言語処理を行っている脳相当になるのかわからない。
  - 人間の創造性と言語  
現状のコンピュータには創造性が欠けている<sup>23</sup>。例えば、会話、作文、作句、作曲、作画、詰め将棋の問題作成等々である。例えば、俳句は50音17文字=10<sup>29</sup>の組み合わせになるが、日本語になる組み合わせとなる確率は極めて低い。しかし、人間の脳は、天文学的な組み合わせから極めて限られた組み合わせを発見し、かつ、それまでになかった新しいものを作ることができる。これはなぜか？その鍵は言語の文法にあるのではないか。新しい発話を生み出し、同時に言語でないものは脳内で捨てている。「(生み出す)かつ(捨てる)」ことが人の言語処理の特徴である。例えば、俳句、将棋にもルールがあって、極めて多くのパターンを生み出すが、検討に値しないものは捨てることができる。
  - チョムスキーの生成文法  
60年前に提案された。  
句構造の例:  
「みにくい、アヒルの子」  
「みにくいアヒルの、子」  
文節の区切りが本質的で、「間」が言語的な意味もっている。意味が決まれば構造は一意的に決まる。「赤いリンゴ」と「リンゴの 赤」は違う。日本語では右側に重要な情報を持ってくる。英語は最初のほうが重要となる。日本語は最後を聞くまで結論がわからない。  
文法の規則からいくらでも長い文章をつくるのが原理的に可能である。
  - 人間の言語の特徴  
人の言語には「再帰的」という特徴があり、無限を取り扱うことができる。言語の構造を木構造で表すことができる。文法的操作(意味的なものは除く)は2つのオペレーション「併合」と「サーチ」からなる。文法処理に係るニューラルネットワークを解析していくことで、人間特有の知の構造が分かってくるのではないかと
- ### (2) 文法処理に関わる脳活動を探る
- 実験の目的  
脳の文法(句構造文法)の処理に特化している部位を特定すること。  
句構造文法では埋め込み文の深さに比例して処理にかかる脳の負荷が上がる(血流量が増える)という結果を予測し、負荷が異なるタスク間の脳の活動の差分を取ることで、文法処理と語彙処理の役割を担っている部位を特定する。
  - 実験方法

意味処理による脳の活動を除外するために、無意味な造語に助詞をつけて文を作り、それらの助詞を手掛かりに構文解析を行わせるタスクを、文法処理が異なる埋め込み文、単文、連文に対応して3種類用意する。

また、語順による処理の影響を排除するため、助詞をつけない造語の単語をマッチングさせるタスクを逆順文字列、同順文字列と2種類用意する。

それぞれのタスクでは、4単語からなる短文と、6単語からなる長文のいずれかが用いられる。

- 文法解析の負荷の指標として、木の深さを「併合度(DoM, Degree of Merger)」とし、構文の特徴の検索を行う回数を「サーチ数」とする。

各文の併合度及びサーチ数を以下に記す。

長い埋め込み文:埋め込み文(長)

併合度=5, サーチ数=3

短い埋め込み文:埋め込み文(短)

併合度=3, サーチ数=2

長い単文:単文(長) 併合度=3, サーチ数=2

短い単文:単文(短) 併合度=2, サーチ数=1

長い連文:連文(長) 併合度=2, サーチ数=3

短い連文:連文(短) 併合度=2, サーチ数=2

- 実験結果1: 文法処理に選択的な活動  
埋め込み文での文法処理の信号変化と語順処理による信号変化を比較することで文法を処理する部位を特定した。

逆順文字列と同順文字列双方ともに 併合度=1, サーチ数=(短:2, 長:3)

Nested'' = 埋め込み文 - 単文

Reverse'' = 逆順文字列 - 同順文字列

Nested'' > Reverse''となる脳の信号の変化を調べると左下前頭回と左縁上回での変化が大きいことがわかり、文法処理を担っていると考えることができる。

- 実験結果2: 併合度に選択的な言語野の活動変化  
以下の差分を定義し、脳の信号変化量と比較を行う。

Nested'(長) = 埋め込み文(長) - 連文(長)

Nested'(短) = 埋め込み文(短) - 連文(短)

Simple'(長) = 単文(長) - 連文(長)

Simple'(短) = 単文(短) - 連文(短)

上式による併合度の差分は、順に 3, 1, 1, 0 となり、これらの差分は左下前頭回の信号変化量に近い値になっている。

一方、(併合度 + サーチ数)の差分は、順に 3, 1, 0, -1 となり、左縁上回の信号変化量に近い値になった。

左下前頭回の活動の差分は併合度の理論上の計算の差分と対応しており、併合度つまり構文解析に選択的な活動変化が見られた。

一方、左縁上回の活動変化は併合度+サーチの差分の計算と一致し、文法処理に加え単語の検索も行っていると思われる。

- 実験結果3: 左下前頭回と左縁上回の間関係  
一方、DCM(Dynamic Causal Modeling)の計算によると、左下前頭回から左縁上回への活動の関係が正の相関

(+0.22)になっており、文法処理の活動は左下前頭回から左縁上回へトップダウンに伝えられることがわかる。以上の結果により、左下前頭回が文法処理に特化していることがわかった。

### 5.3 質疑応答

- Q: 言語の木構造は、先天的なものか、後天的なものか  
A: 人間は、環境にある言語を、文法を教えずとも習得して、2～3才で構造のある文を自然にしゃべる。また、これは文法の構造を教えることができるのではない。先天的に文法のひな形が脳に何らかの形で埋め込まれていると考えられる。
- Q: チョムスキーの生成文法が生得性のものとして、その進化の過程はどうなっているのか？なぜ人以外の動物ではだめなのか？  
A: 人間の文法の獲得が、進化の過程で出来たものであることは間違いない。遺伝子のレベルでは科学的に解析できるかもしれないが、人間に至る過程の動物(ネアンデルタール人等)は現存しないので、実証研究はできない。また、動物は人間のような文法をもっていない。
- Q コンピュータは人の脳に比べてオーバースペック過ぎて、人の言語を写し取れないのでは、というのはどういうことか？  
A: チューリング・マシンは確かに万能だが、人間の言語を持つ計算上の制約を適切に与えない限り、言語以外の記号列を大量に生成してしまう。たとえ膨大なテキスト・データをコンピュータに入力したとしても、文法的な新しい文を限りなく創り出す人間の能力を再現することはできない。それは原理的に不可能という意味ではなく、文法計算のアルゴリズムが十分に解明されれば、ある程度まで自然言語に近づくことはできるかもしれない。

### 6. 全体討論

- Q: マルチモーダル、カテゴリゼーションでは、アフォーダンスの取り扱いは？  
谷口氏: 結局はある種の関連性をデータセットからどう見いだすかという話だが、道具を見せてその操作を見せる、そして、道具を使った結果を見せるというのを見せて、データの関連性を見ていくことになる。アフォーダンスの概念も解釈は多様だが、構成論的なアプローチで組む事はできると思う。
- Q: 今後のロードマップについて:「今後の人間のような言語とか記号を取り扱う人工知能を扱う場合のハードルは？」  
林氏: テーマが大きい、視覚研究の分野からするとスタティックな入力では学習させているが、時間情報を含めた場合に、概念そのものを獲得していくのが大きな課題である。言語では、繰り返しの時間的な構造があるが、そのことと視覚がどう結びついていくか、新しいアイデアがあれば、脳の全体的な理解や、人工知能システムの実現が開けていくのではないと思う。
- 谷口氏: 概念とか、言語とか、言葉自体がとても曖昧なので、多くの人が記号論理的な意味での「記号」をナイーブに考えていて、カテゴリに対するラベルとしてしか見ていない。コミュニケーションにおいて、記号やその解釈は文脈において、変わるし、作りたい放題でもある。記号とか、言語、概念をどのようにマルチモーダルや社会的な意味やインタラクションも含めて、定式化していくのか、記号とか概念とかの言葉の意味が有る程度踏み込むといつもわからなくなる、シンタックス(文法)はチョムスキーのおかげである程度できていて、それで議論が展開できた経緯がある。より構成論的に、言語的な刺激だけでなく、マルチモーダルなものとして記号や

概念を構築していく必要があるのではない、それが進む事で記号や概念の何が問題なのかが見えてくるのではないかと思う。「概念とはなにか?」「記号とはなにか?」自体が問題として掲げられても、曖昧な定義のものを曖昧に扱って、何が問題かわからなくなるのが問題なのではないかと思う。

酒井氏: ニューラルネットの場合は、言語計算に必要なシンボル処理を扱えない。扱う構造がないし、集団的な活動パターンからシンボルをどう対応づけていくかが一番難しい。これはニューラルネットが出た当初からの難問であり、まだ、そこには突破口がない状態である。

今回の発表にもあったように様々な画像のパターンやカテゴリをマップできる可能性からみれば、言語も同様にできると思いがちだが、もしかすると、我々の脳は言語をそのまま扱ってはいないのではないかと思う。つまり、「脳言語」が存在するだけで、それを脳内で日本語や英語に翻訳して出力しているのかもしれない。そこが見えないので、まだ適切な脳の計算モデルが作れない。チョムスキーに言わせると、それが「普遍文法」というもので、どの言語にも共通したものがある。だから、赤ちゃんは人間として生まれてくれば、何語でも話せるようになるのではないか。そのように理解すると、それぞれの言語の個別性ではなく、普遍的な原理だけが脳に組み込まれていて、後は環境に合わせてパラメータを合せるのが、言語獲得だといえる。それを考えたときに、我々が日常使っている言語そのものにマップするという方法では表面的すぎると考える。

### 7. まとめ

視覚情報の脳内情報処理、ディープニューラルネットワークによる教師付き学習による機械学習のあるレイヤーと脳内の高次処理をしている信号活動に相関がみられる。見ている状態の脳の信号を取り出して、コンピュータに入力し、見ているものを認識できるように復号可能となってきている。

記号創発、人工知能の身体性として、ロボットでいうところの「モーター・センサー」系に相当するものだとすると、人が身体と脳の組み合わせで創発した記号＝言語とは、全く別の記号＝言語になる可能性がある。それを人間は理解できるのか考える必要がある。

人の脳の特徴は、文法を理解できることにある。無限という概念もその文法に織り込まれた再帰的な処理が可能であることによる。進化過程でそうなったかも知れないが、他の動物との言語能力の断絶は大きく、また、進化の過程は実験的には証明できない。しかし文法の計算が人間の知性のあり方とすれば、汎用人工知能においてもそれを何らかの形で組み込む必要がある。

### 謝辞

本会の実施にあたり、会場をご提供いただいた株式会社リクルートテクノロジーズ様およびニコ生放送をいただいた株式会社ドワンゴ様に感謝いたします。



## 参考文献

- <sup>1</sup>山川, 市瀬, 汎用人工知能輪読会の発足とその後の活動, 人工知能, 29, 3, 2014.  
[http://www.ai-gakkai.or.jp/wp/wp-content/themes/shinra-of-the-sun/pdf/265-267\\_293.pdf](http://www.ai-gakkai.or.jp/wp/wp-content/themes/shinra-of-the-sun/pdf/265-267_293.pdf)
- <sup>2</sup>荒川, 山川, 市瀬, 汎用人工知能の研究動向, 人工知能学会全国大会, 2014.  
<https://kaigi.org/jsai/webprogram/2014/pdf/174.pdf>
- <sup>3</sup>清嶋, 総務省が「2045年の人工知能」「シンギュラリティ」の研究會, 第1回會合で激論, 日経コンピュータ, 2015.  
<http://itpro.nikkeibp.co.jp/atcl/news/15/020900464/>
- <sup>4</sup>S. Harnad, The Symbol Grounding Problem, 1990.  
[http://www.summer10.isc.uqam.ca/page/docs/readings/HARNAD\\_Stevan/symgro.pdf](http://www.summer10.isc.uqam.ca/page/docs/readings/HARNAD_Stevan/symgro.pdf)
- <sup>5</sup>山川, ドワンゴ人工知能研究所 設立趣旨, 2014.  
<http://ailab.dwango.co.jp>
- <sup>6</sup>小松, 視覚の解説, 総合研究大学院大学生理学研究所小松研究室, 2008.  
<http://www.nips.ac.jp/scinfo/kaisetsu.htm>
- <sup>7</sup>中山, Deep Learning と画像認識～歴史・理論・実践～, 東京大学大学院 中山研究室, 2014.  
[http://www.slideshare.net/nlab\\_utokyo/deep-learning-40959442](http://www.slideshare.net/nlab_utokyo/deep-learning-40959442)
- <sup>8</sup>NVIDIA JAPAN, 1003: Deep Learning によるビッグデータ解析～手法や CUDA による高速化, GTC Japan, 2014.  
<https://www.youtube.com/watch?v=whZgLPjE0UM>
- <sup>9</sup>浅川, [脳に挑む人工知能 3] 脳科学とのコラボで「人を超える知性」を目指す, ITpro Report, 日経コンピュータ, 2014.  
<http://itpro.nikkeibp.co.jp/atcl/column/14/090100053/092200012/>
- <sup>10</sup>R. Hayashi, S. Nishimoto, “Decoding visual information in monkey IT cortex using deep neural network”, Proceedings of Life Engineering Symposium 2013 (LE2013), pp.511-514, 2013
- <sup>11</sup>西本, エンコーディングモデルを用いた視覚情報処理研究: 情報表現, 予測, デコーディング, 2012.  
[https://www.jstage.jst.go.jp/article/jnns/19/1/19\\_39/\\_pdf](https://www.jstage.jst.go.jp/article/jnns/19/1/19_39/_pdf)
- <sup>12</sup>日本語 WordNet, 情報通信研究機構 NICT 情報分析研究室, 2006-2012.  
<http://nlpwww.nict.go.jp/wn-ja/>
- <sup>13</sup>C. Szegedy, Going Deeper With Convolutions, GoogLeNet team, 2014.  
[https://www.youtube.com/watch?v=ySrj\\_G5gHWI](https://www.youtube.com/watch?v=ySrj_G5gHWI)
- <sup>14</sup>林, 視覚野の神経情報を用いたロボットハンド制御技術の開発, 立石科学技術振興財団 助成研究成果集(第23号), 2014.  
<http://www.tateisi-f.org/html/reports/h25reports/2031013.pdf>
- <sup>15</sup>伊藤, 「意味のない」脳科学のススメ, 日本神経回路学会誌 Vol.21, No.2, 43-49, 2014.  
[https://www.jstage.jst.go.jp/article/jnns/21/2/21\\_43/\\_pdf](https://www.jstage.jst.go.jp/article/jnns/21/2/21_43/_pdf)
- <sup>16</sup>谷口, 記号創発ロボティクス 知能のメカニズム入門, 講談社, 2014.
- <sup>17</sup>谷口, コミュニケーションするロボットは創れるかー記号創発システムへの構成論的アプローチ, エヌティティ出版, 2010.
- <sup>18</sup>谷口ら, 記号創発ロボティクスとマルチモーダルセマンティックインタラクション, 人工知能学会, 2011.  
<https://kaigi.org/jsai/webprogram/2011/pdf/395.pdf>
- <sup>19</sup>Rodney A. Brooks, A Robot Walks; Emergent Behaviors from a Carefully Evolved Network, A. I. Memo 1091, 1989  
<http://people.csail.mit.edu/brooks/papers/AIM-1091.pdf>
- <sup>20</sup>酒井, 言語の脳科学: 脳はどのようにことばを生み出すか, 中公新書, 2002.
- <sup>21</sup>S. Ohta et.al., Syntactic computation in the human brain: The degree of merger as a key factor, PLOS ONE 8, e56230, 1-16, 2013.  
<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0056230>
- <sup>22</sup>成田, 飯島, 酒井, 人間以外の動物に「文法」は使えるのか?, Brain and Nerve 66, 276-279, 2014.  
[http://mind.c.u-tokyo.ac.jp/Sakai\\_Lab\\_files/Staff/KLS\\_PaperJ/KLS2014Jd.pdf](http://mind.c.u-tokyo.ac.jp/Sakai_Lab_files/Staff/KLS_PaperJ/KLS2014Jd.pdf)
- <sup>23</sup>酒井(編), 芸術を創る脳: 美・言語・人間性をめぐる対話, 東京大学出版会, 2013.