

# 人間の記号化機能の一考察

## A study of the human functions for the symbolization

岡谷 基弘<sup>1</sup>

Motohiro Okaya<sup>1</sup>

<sup>1</sup>フリー

<sup>1</sup>Freelance

**Abstract:** How does human symbolize the world? I introduce “Double Articulation multi-Dimensional Symbolization that are clustered and reduced into Stories as the world model” (DAMDiSS). Humans get a lot of “sensor data” via their sensory organs. The “raw” data are clustered in an unsupervised manner. And then, I assume double articulation structures in all modal data. The “meaning” is a concept which is consist of several modal clustered “raw” data. I also introduce time-series “meanings” as “Story”. A “big Story” is consist of several “Stories”. Humans can make a desirable “big Story” by selecting various kinds of “Stories” in their memories and can modify the “big Story” after acquiring new “raw” data. This process is just the “Bayesian inference algorithm” itself. The consciousness is a “Bayesian inference algorithm” which enables us to form and modify the optimal plan by considering year-order future value.

## 1. 背景

近年 AGI(Artificial General Intelligence)の実現に対する期待が高まりつつある。そのため、各種の分野で精力的に研究が行われているが、人間がどのように世界を記号化しているかという点に関して、統一的な見方は得られていない。そこで、本論文では人間の記号化機能についての一考察を述べる。

## 2. 人間知能の機能

人間には大きな3つの根本的機能があると考えられる[1]。1つ目は、「世界の記号化」、2つ目は「望ましい世界を求め続ける広義の欲求」、3つ目は「予測に基づく方策決定」である。このうち、「世界の記号化」について本論文で詳しく論じる。

### 2.1. 時系列刺激の多次元記号仮説

人間は生まれてから様々な刺激を受ける。これは、視覚を通じた画像の時系列的刺激や、聴覚を通じた音響の時系列的刺激、皮膚を通じた圧力・温度の時系列的刺激、内受容感覚による種々の時系列的刺激などである。人間は、生まれてから「一定のルール」に基づいて、これらの時系列刺激を教師なしクラスタリングしていると推定している。記号創発ロボテ

イクスの分野では、ロボットに画像や触覚アレイセンサの情報をクラスタリングさせることで概念形成[2]を実現している。人間は、現在のロボットよりもはるかに多くの画像情報、触覚情報等を扱えるが、それらの情報量を適宜削減しクラスタリングすることで概念を形成しているのである。

ここで、概念形成について一つの仮説を提案する。人間が受け取る時系列刺激は、脳内ではニューロンの発火パターンとなっているが、この「流れ方」は各感覚器官を次元（またはそれをさらに抽象化した次元）とする多次元空間の超立体として表現可能であるという仮説である。図1は、時系列データが多次元の超立体として表現される概念図を表している。人間がある言葉に対して感じる「意味」とは、この多次元の超立体を参照していると考えられる。例えば、「りんご」という言葉の意味は、赤くて甘い果物で、噛むとシャリシャリして美味しく、空気中で放っておくとだんだん色が変わるといったものだが、これは、視覚刺激、触覚刺激、味覚刺激等を次元とする超立体で表現されると考える。果物の「なし」は「りんご」と似た超立体になるが、視覚刺激軸、味覚刺激軸の値(カテゴリ)が異なるといった具合である。

この仮説を用いると、ロボットの物体概念形成において見通しが良くなる。つまり、各センサー情報を元にした次元を使用し、物体概念をテンソルで表現できる。さらに具体的にいうと、プログラミング言語の Python の numpy でテンソルを作り、その中

の値を決めることで「意味」を定義できる。

人間が世界を記号化する際は、「自分自身がある世界の対象をどう記号化しているか」自体も記号化の対象となる。この自己認識の記号化機能は、新たな次元を追加し、意味の多次元超立体のその次元への射影を考えることで表現できるのではないかと考えている。

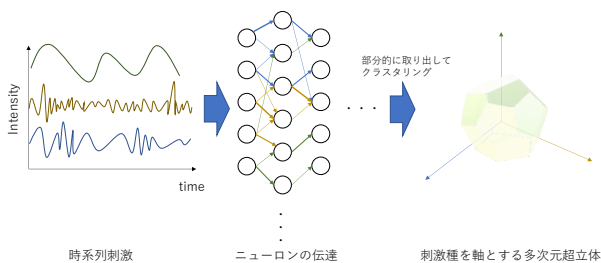


図 1 時系列データの脳内処理と多次元記号仮説

## 2.2. 多次元記号仮説とマクロな世界モデル

前節で述べた多次元超立体による意味の定義を行うためには、実世界をセンシングした「生データ」に対して何らかの処理を行う必要がある。以下に、その具体的な手順の提案と課題を述べる。

音声の認識を例にとって考察する。図 2 に音声の認識における生データの認識の概要を示す。初めに横軸が時間で縦軸が何らかの強度となるデータが得られ、そこから特徴量表現の時系列データを得る。ここで、二重分節構造[2]を想定する。特徴量表現の時系列データは音声であれば「音素」と呼ばれる。例えば、「w”や”a”である。「音素」には人間は意味を与えていないが、その組み合わせに対して意味を与えている。例えば、「watashi::わたし」などである。

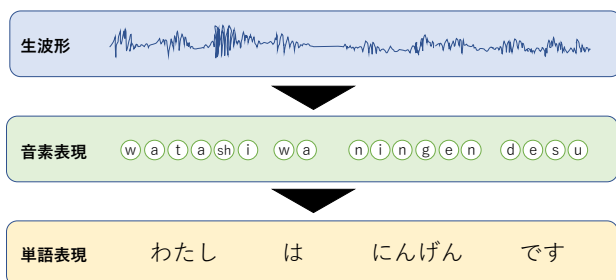


図 2 音声データ処理概念図

「音素」は有限だが、組み合わせは極めて大きな数になる（50 個の音素を 3 つ組み合わせる通りの数は  $50 \times 49 \times 48 \approx 10$  万）ことから、多数の「意味」を認識できるようにするため、このような二重分節構造により認識を行っていると考えられる。

ここで、視覚や触覚データ等も二重分節構造を取っていると仮定する。「音素」に相当する、組み合わせにより意味を定義するそれ自体は意味を持たない単位を「モーダル素」と呼称することにする。「モーダル素」を生データから生成する上で以下の課題がある。

- 課題①：連続的な生データをどこで分節するかを教師なし学習する必要がある
- 課題②：時間的な伸縮にロバストな学習を行う必要がある。

人間の知覚は常に連続データを取得しているが、「どこで区切ってモーダル素を構成するか」をデータのみから学習する必要がある。ここで、モーダル素構成の特徴として、多少の時間伸縮があっても同じものと認識することを考慮する。これは、図 3 に示すように、「a」という音素が 0.4 秒の発音でも 0.5 秒の発音でも同じ「a」と認識されることが多いであろうという直感に基づいている。こうした多次元時系列データの教師なし分節化というテーマでは、日本語の発話音声データに関して言語モデルと音声モデルを音素の持続時間をも含めて予測する HDP-HLM[4]手法が提案されており、この手法が他のモーダルにも有効であれば、モーダル素の教師なし分節化は可能と考える。

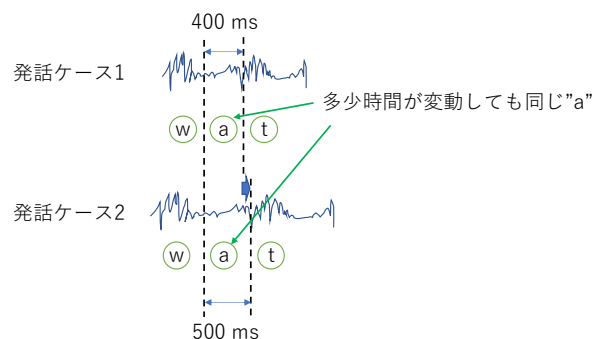


図 3 音素認識の時間変動のロバスト性

次に、適当なモーダル素を得ることができたという前提で考察を続ける。ここで、前述の「意味」の超立体を構成する軸のカテゴリはモーダル素で構成されると考える。こう考えることで、「意味」を生データからの変換として一貫して捉えることができる。

ここで、超立体を構成する課題として以下が考えられる。

課題③：一つの意味を構成するモーダル素をどこで分節するかを教師なし学習する必要がある

課題④：上記について時間的な伸縮にロバストな学習を行う必要がある。

課題③の概念図を図4に示す。各モーダルの「生データ」からモーダルごとのモーダル素の時系列表現が得られたとして、複数のモーダル素を「いつ」分節するかを教師なしで学習する必要がある。また、これについても直感的には、多少の時間伸縮にロバストに分節される必要があると考える。

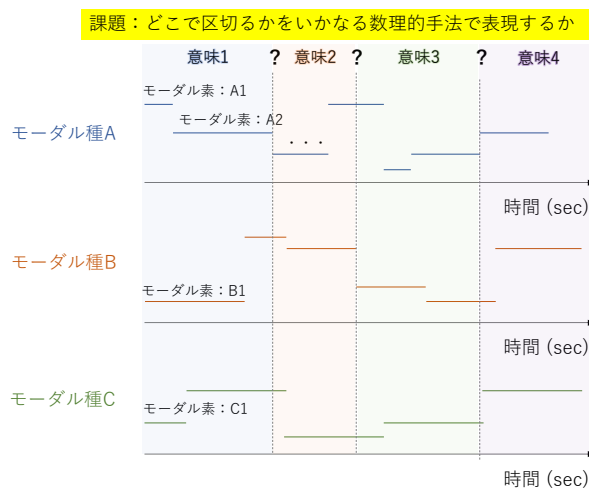


図4 意味を構成する複数のモーダル素の区切り方の課題

さらに、図5に示すように、多次元モーダル素超立体が時間的に少しずつ変化するという形で、人間は意味を認識していると考えられる。この「意味」の時系列的連続の一区切りを「ストーリー」<sup>1</sup>と呼ぶことにする。

“STORY”：ひとまとまりの意味の時間的連続

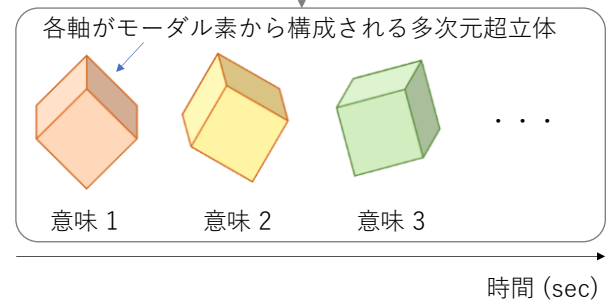


図5 意味の時系列認識:ストーリー

「ストーリー」をさらに多数組み合わせ、時間的に縮約した表現を得ることで、分、時間、日、年といったカテゴリとしての時間軸を持った「大ストーリー」を作ることができる（図6）。この「大ストーリー」は、例えば、任意の個数の「ストーリー」をVAE(Variational Auto-Encoder)のような仕組みにより圧縮し構成できると推定している。そして、必要に応じて記憶の中から“decode”し若干ぼやけた感覚刺激の再現を得るのではないかと考えている。この「大ストーリー」の構成は過去の経験の縮約表現であるが、「ストーリー」を任意に組み合わせることで、予測としての「大ストーリー」を得ることもできるのではないかと。つまり、人間の世界モデルとは、「ストーリー」の組み立てによって構成されていると考える。

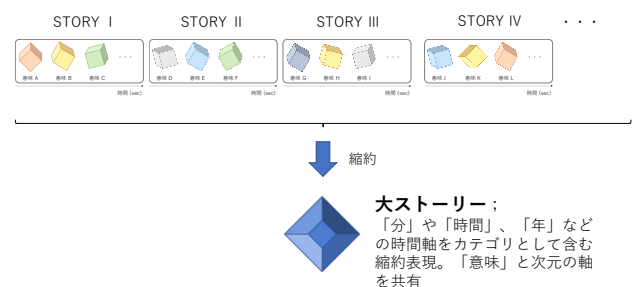


図6 大ストーリーの構成

世界モデルおよび予測を「ストーリー」の組み立てによって捉えることで、人間の事実上無限と感じられる自由意志による方策選択について、これを数理的な問題へ変換できる可能性が見えてくる。つま

<sup>1</sup> 「意味」や「ストーリー」には自然言語による記号が与えられることがある。例えば、皮膚が高温を感じたとき、「熱い」と感じるが、これは皮膚感覚刺激のモーダル素が複数組み合わせられたものである。また、「熱い風呂に入って気持ちいい」というものはストーリーである。皮膚感覚

刺激としての熱さや温水による圧力刺激、呼吸の若干の苦しさなどが合わさったものである。しかし、全ての「意味」や「ストーリー」に自然言語が割り当てられているわけではない。

り、図7に示すように、まず価値 MAX の大ストーリーを構想し、その大ストーリーに至る道筋を記憶の中にある全てのストーリーの組み合わせにより構成し、その結果として行動が行われると考える。これにより、方策の選択を「組み合わせ最適化問題」として捉えることができると考えられる。この考え方における課題を下記に示す。

- 課題⑤：価値が高まる「大ストーリー」の具体的な構想方法（時間軸の決定、「大ストーリー」を構成する「ストーリー」の選択と組み立て）
- 課題⑥：価値が高まる「大ストーリー」に現在から向かう道筋としての一連の「ストーリー」の最適配置
- 課題⑦：未来の価値の現在価値への割戻し方法
- 課題⑧：少なくとも数万以上あると予想されるストーリーの組み合わせ最適化を現実的な時間で計算するアルゴリズムとハードの開発

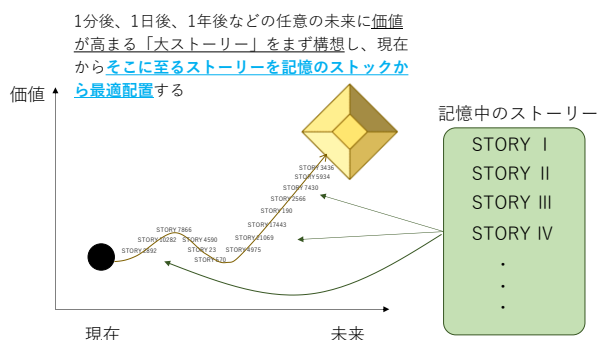


図 7 組み合わせ最適化問題としての方策決定

### 2.3. 意識とは何か

前節までの議論をもとに「意識」とは何か考察する。「意識」とは何かという問に対し、人々は様々な解釈をしてきたが、本節では見方を変えて、「意識は何を実現しているか」という視点で考察する。

筆者の仮説は「意識は年以上のオーダーでの最適方策の立案とその方策の微調整を実現する機能」であるというものである。仮に、1年後に何らかの精神的報酬を得るための方策を考え、今行動を開始したとする。1年後までの方策にはざっくりと三ヶ月後や半年後の状態の計画があるはずである。ここで、ちょうど半年経ったとき、計画の変更または調整の必要性を吟味したいと思ったとする。計画の変更のためには、半年前に想定した今（半年後）の状態と、今現在の状態を比較し、半年後の望ましい状態に至

る方策を再度考える必要がある。ここで、仮に意識がなかったとすると、今現在の状態を客観的に認識できない。快不快は認知できたとしても、それを予想した快不快と比較するためには、快不快のメタ認知が必要である。これを実現しているのが意識だと考える。ただし、年以上のオーダーの計画立案機構があるならば必ず意識があるということではない。

前節までで、感覚器官で取得した「生データ」から「モーダル素」、「モーダル素」多数から多次元超立体としての「意味」、「意味」の時系列的連続から「ストーリー」、「ストーリー」を多数組み合わせた「大ストーリー」という形で人間の記号化機能を論じた。意識とは、より大きな報酬が得られる年オーダーの後の理想状態に至る道筋を、記憶の中にある「ストーリー」や「大ストーリー」の組み合わせにより反実仮想的に構成する仕組みと考える。この反実仮想的な「理想状態に向かうストーリー」はいわば「事前分布」であり、経験した出来事（感覚器官によるセンサーデータ）により更新され「事後分布」を得るベイズ推論が意識の本質であるという仮説を提案する（図8）。

計算時間の問題はあるものの、上記の機能は「肉体」を持つロボットに実装することは理論的には可能と思われる。仮にこの機能をロボットに実装し、十分に学習させることができたとしても、そのロボットが「本当に」意識を保有できたか否かは証明不可能だと考えられるものの、そのロボットの行動様式を人間が観察すると、人間は「ロボットに意識がある」と感じるかもしれない。

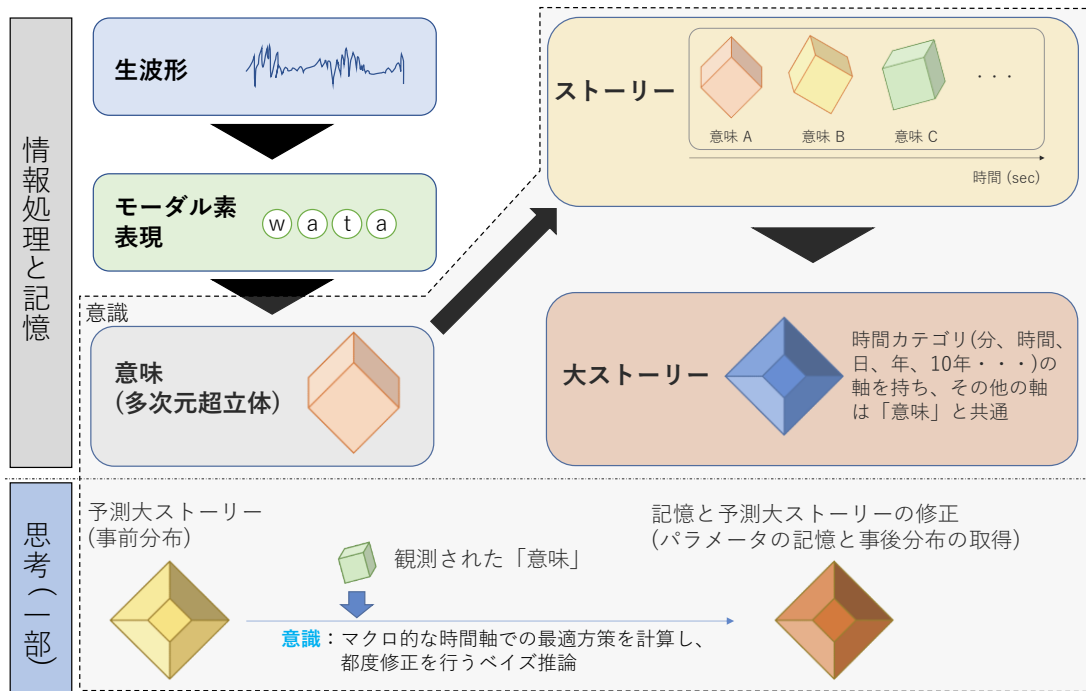


図 8 感覚器官から得た生データから意味を得る仕組みと意識の概念図

### 3. おわりに

人間の記号化機能の考察を行った。感覚器官で取得した「生データ」から「モーダル素」、「モーダル素」多数から多次元超立体としての「意味」、「意味」の時系列的連続から「ストーリー」、「ストーリー」を多数組み合わせ縮約した「大ストーリー」を得るという記号化機能仮説を提案した。予測として望ましい「大ストーリー」を構想し、方策としてそれに至る記憶の中の「ストーリー」の組み立てを考えると、これは組み合わせ最適化問題として表現できる可能性がある。「意識」とは「年以上のオーダーでの最適方策の立案とその方策の微調整」を実現するベイズ推論であるという提案を行った。我々人間の意識の一部は上記機能の発露として得られていると考える。

上記仮説を検証するためには、実世界でロボットに上記機能を実装し、年単位で時間をかけて学習を行うか、それをシミュレーション空間上で行う必要がある。いずれにしても、意味の計算は膨大な時間がかかり、組み合わせ最適化問題として方策計算を行うためには汎用量子コンピュータが必要となる可能性があるものの、単純化したモデルで検証を行っていく予定である。

### 参考文献

- [1] 岡谷基弘: AGIのデザインに向けた人間機能の一考察、人工知能学会研究会資料、SIG-AGI-013-03、(2019)
- [2] 谷口忠大, 記号創発ロボティクス 知能のメカニズム入門, 講談社選書メチエ, (2014)
- [3] Tadahiro Taniguchi, et al, “Nonparametric Bayesian Double Articulation Analyzer for Direct Language Acquisition From Continuous Speech Signals”, IEEE Transactions on Cognitive and Developmental Systems, VOL. 8, No. 3, (2016)