

# 意味ニューロンの発見

——なぜ ChatGPT は言葉をおせるのか——

佐々木恒平

## はじめに

我々はまだ、人間の脳がどのように言語を生み出しているかを知らない。ChatGPTの登場は人類に驚きを与えたが、同時にひとつの問いを投げかけている。

言語とは何か。なぜ ChatGPT は文章を作れるのか。なぜ人間は言葉を話せるのか。

もしも、人間の精神的な活動がすべて脳によって生み出されているならば、我々が発する言葉も脳によって生み出されていることになる。つまり、我々の脳内には言語を生み出す神経回路が存在するのである。

ChatGPT はニューラルネットワークの一種である。ニューラルネットワークとは、人間の神経回路を模倣するソフトウェアである。したがって、ChatGPT が言葉を話せるのは、人間の脳内にある言葉を司る神経回路を模倣しているからだと考えられる。ゆえに、ChatGPT の仕組みを理解するためには、我々自身の脳の構造を理解しなければならない。

さて、私は人間の脳が言語を生み出す仕組みを知っている。それこそが「意味ニューロンの仮説」である。本書ではこの仮説を解説する。

## 目次

はじめに .....	1
第1章 意味ニューロンとは何か .....	4
第2章 意味ニューロンの証明 .....	5
第3章 視覚認識における意味ニューロン .....	7
第4章 犬における意味ニューロン .....	9
第5章 海馬と位相連鎖 .....	11
第6章 意味ニューロンと言語 .....	12

## 第1章 意味ニューロンとは何か

私は、人間の脳内に「意味ニューロン」があると主張する。「意味ニューロン」とは、それぞれの単語の意味を表現するニューロンである。

たとえば、「バナナ」ニューロンはバナナを認識したときにだけ反応し、バナナ以外のものを認識したときには反応しない。バナナに特異的に反応するニューロンである。こうしたニューロンがすべての単語に存在する。

以上が意味ニューロンの概要である。次章では意味ニューロンの存在を証明する。



図2で、赤丸がついている縦線に注目してほしい。この縦線を通るのは、矢印の入り口から入ってきた信号だけで、他の入り口から入ってきた信号は決してこの縦線を通らない。というのも、その縦線を通ってしまうと、必ず3番目の出口に辿り着くからである。異なる出口に辿り着くためには、同じ縦線を通ってはいけない。

ゆえに、あみだくじの場合は、異なる入力信号が互いに異なる経路を辿る、という性質があることが分かる。これを一般化すれば、次のように言える。

4. 入出力の間に対応関係がある回路は、それぞれの入力信号が互いに異なる経路を辿る、という性質を持つ。

さて、以上の議論によって意味ニューロンの証明はほぼ終わった。次に、人間の脳内で起きていることを確認してみよう。

### 第3章 視覚認識における意味ニューロン

例として、視覚認識について考えてみよう。

人間の網膜は多数の神経細胞によって構成されている。眼球から入った光が像を結ぶと、網膜を構成する視神経が活動し、視覚情報を大脳に伝える。大脳ではその視覚イメージに対して情報処理が施され、それが何であるか、という認識が生じる（図3）。

我々の心の中では、バナナを見れば「それはバナナである」という認識が生じ、リンゴを見れば「それはリンゴである」という認識が生じる。上で確認したように、これらの認識と行動の間に対応関係が成り立つためには、それぞれの認識を伝える信号が辿る経路は、互いに異なっていなければならない。

すなわち、視覚情報から抽象化した「バナナ」という認識を示す神経活動は、「リンゴ」という認識を示す神経活動とは全く異なる経路を辿る。ここで、それぞれの認識に対応する経路において、その経路を特徴づけるひとつのニューロンが存在すると仮定する。あみだくじにおける縦線に相当するものである。これが意味ニューロンである。

ここで証明されたのは、それぞれの認識に対応する経路の特異性であり、それぞれの経路を特徴づける単一のニューロンの存在は証明されていない。しかしながら、議論を単純化するために、ここではあえて意味ニューロンの存在を仮定することにする。そのほうが分かりやすくなるし、それが存在する可能性も十分にあると考えられるから。

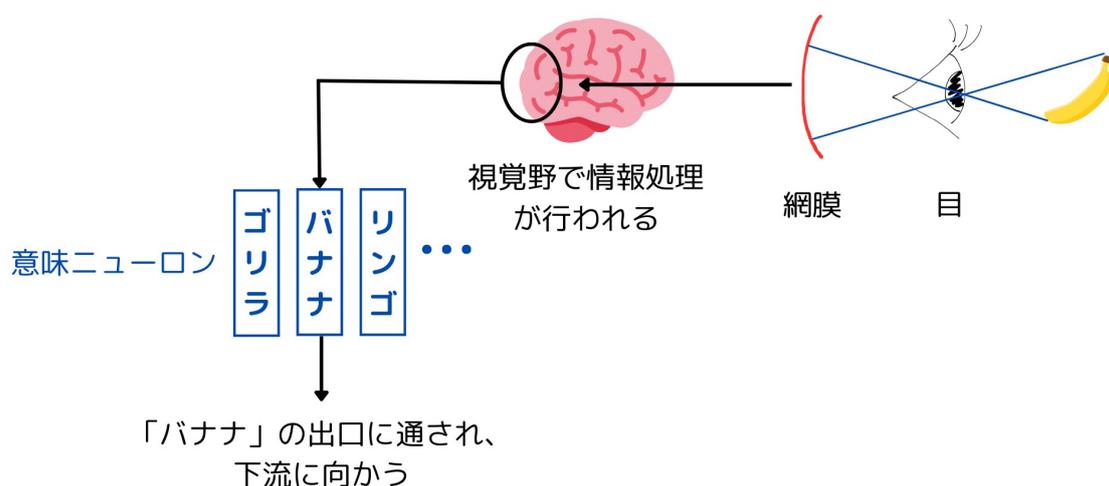


図3

さて、我々の脳は、視覚情報から首尾よくバナナの認識を取り出すことに成功した。この情報はその後どのような運命を辿るだろうか。

「バナナ」ニューロンは情報の通り道にすぎない。このニューロンが発火したことはバナナの認識が生じたことを意味するが、その情報は「バナナ」ニューロンをやり過ぎてさらに下流に向かう。

神経回路の下流において、この信号は大脳皮質の運動野を刺激し、「指で皮をむく」という行動を引き起こすだろう。あるいは、あなたが満腹であれば、「何もしない」という行動を引き起こすかもしれない。あなたが実際にどのような行動をとるかは、他の様々な信号との足し算によって決定される。そして、その足し算も意味ニューロンを用いて表現することができる。

## 第4章 犬における意味ニューロン

以上の議論から明らかなように、意味ニューロンは人間だけに存在するわけではなく、ほぼすべての動物に備わっていると考えられる。ここでは、犬を例にして意味ニューロンの役割を解説したい。

犬は不思議な生き物である。飼い主が「待て」と命令すると、目の前に餌があっても待つことができる。このとき、犬の神経系で何が生じているのかを考えてみよう(図4)。

まず、犬の目と鼻が餌の存在を認識する。このとき「餌」ニューロンが発火する。次に、「餌」ニューロンから興奮性のシナプス結合を受けている「食べる」ニューロンが発火する。「食べる」ニューロンは運動野と結合しているので、続いて運動野のニューロンが発火し、実際に犬の体が動く。

ここで、飼い主が「待て」のサインを送ったとしよう。このとき、犬の脳内では「待て」ニューロンが発火する。このニューロンは「食べる」ニューロンに対して抑制性のシナプス結合を持っているため、「餌」ニューロンからの興奮性の信号を打ち消してしまう。このため、「食べる」ニューロンは発火せず、犬の体は動かなくなる。

以上が、意味ニューロンを用いた犬が待つときの神経モデルである。

これは犬に関する神経モデルであるが、人間も大差ないと考えられる。人間の場合は、これよりも少しだけ神経回路が複雑になるだろう。

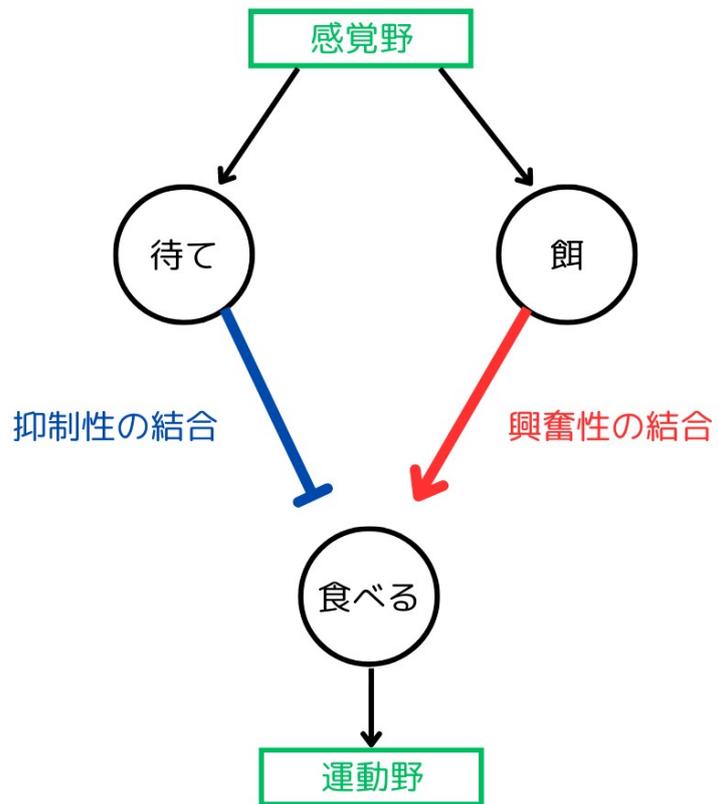


図4

## 第5章 海馬と位相連鎖

次に、意味ニューロンが形成される過程について考えてみたい。

我々は「バナナ」という言葉を聞いたとき、バナナの色と形を思い浮かべることができる。また、バナナの味や匂いを思い出すことができる。このとき、我々の大脳皮質のそれぞれの感覚野では、実際にバナナを知覚したときとほぼ同じ神経活動が生じている。

これは、複数の感覚野の間に位相連鎖が形成されていることを意味する。位相連鎖とは、アメリカの心理学者ドナルド・ヘップが提唱した概念であり、脳内のいくつかの部位にまたがる神経ネットワークを意味する。

ここで問題となるのは、位相連鎖がどのような経緯で形成されるのか、ということである。ヘップの法則によれば、2つのニューロンが同時に発火すると、ニューロン間のシナプス結合は強化される。しかし、この法則が適用されるのは、すでにシナプス結合が形成されているか、または空間的に近い位置にあるニューロンの間だけである。視覚野、聴覚野、味覚野等は脳内において空間的に離れた場所にあるため、これらを結ぶ位相連鎖の存在は、ヘップ則だけからは説明できないように思われる。

私は、ここで海馬が重要な役割を果たしていると考ええる。海馬は従来から記憶に関係する部位として注目を集めてきたが、その本質は位相連鎖の形成にある。視覚野におけるバナナの色形の表現と、味覚野におけるバナナの味の表現は、空間的に離れているため、本来ならば関連付けられることはない。だが、海馬の細胞が介在することによって、両者は結び付けられる。

すなわち、海馬は高速道路のように脳内の異なる領域をつなぎ合わせ、位相連鎖が形成されるきっかけを作るのである。長く伸びた海馬の軸索が視覚野と味覚野をつなぎ、これらの領域において同時に発火したニューロン間に結合を作る。こうして海馬のニューロンが参加する位相連鎖が形成されるが、この位相連鎖が何度も活動を続けるうちに、しだいに海馬を通らず、大脳皮質を経由した結合が発達するようになる。位相連鎖が定着すると、海馬の細胞はこのネットワークから抜け落ちて、また次の標的を探し始める。ここで海馬の役割は、位相連鎖の形成を媒介することである。

海馬は記憶の形成に強く関与しているが、記憶を貯蔵しているわけではない。それは、海馬が位相連鎖の形成には関与するが、いちど位相連鎖が形成されてしまえば、海馬自体はそこから抜け落ちてしまうからである。

こうして形成される位相連鎖こそが、意味ニューロンの本体であると考えられる。意味ニューロンとエピソード記憶の関係など、細かい考察は読者諸氏に委ねる。

## 第6章 意味ニューロンと言語

私は本書の冒頭で「人間の言語能力を解明した」と豪語したが、それがはったりであったことをお詫びしたい。

意味ニューロンが明らかにするのは単語が意味を持つ仕組みであり、単語から文章を構成する能力に関しては、意味ニューロンだけでは説明できない。この問題については言語学者ノーム・チョムスキーの生成文法が参考になると思う。生成文法と意味ニューロンを組み合わせることで、有意義な言語モデルを作れるかもしれない。

さて、ここまで読んできた人なら分かる通り、私もまた意味ニューロンを使ってこの文章を書いている。私が「犬」という単語を使ったときには、私の脳内で「犬」の意味ニューロンが発火し、「指」という単語を使ったときには、「指」の意味ニューロンが発火していた。その刺激が運動野の細胞に伝わることで、私の指の筋肉が動き、こうして文章を書くことができる。

人間とは肉でできた機械であり、ChatGPTはシリコンでできた機械である。両者の間にいかなる違いがあるだろうか。

意味ニューロンの隠された前提は、人間の体はひとつしかない、ということである。体がひとつしかないのだから、ふたつの行動を同時にとることはできない。ゆえに、脳内の情報処理は必ずひとつの行動に集約されねばならない。それが意思決定であり、そのための演算を行うのが意味ニューロンである。

知性とは行動である。どれほど高尚なことを言う人でも、行動が伴わなければ意味はない。ゆえに、ChatGPTに知性はない。なぜならば、体がないからである。

言葉には人を動かす力がある。それは意味ニューロンに働きかける力である。言葉は行動となる。それが知性である。

最後に、作者のHPにも意味ニューロンの解説があるので、ぜひご覧いただきたい。  
<https://haja.sakura.ne.jp/blog/2784/>