# 言語の認知科学第 11 回
# 意味記憶と自己組織化モデル

担当：浅川伸一

2010 年 1 月 13 日

## 1　他者の心を理解するミラーニューロン

ミラーニューロンは，社会的認知と社会的相互作用の複雑な構造に，史上初めて神経生理学的説明を与えるものであると言われている。ミラーニューロンは我々に他人の行動を認識させることにより，その行動の背後にある最も深い動機，すなわち他人の行動の意図を認識させ，理解させているものだとされる。かつては「意図」の実証的な研究はほとんど不可能だと言われていた。意図とはきわめてとらえどころのないメンタルなものであり，実証的な手段ではとうてい調べようがないと考えられてきた。そもそも他人が自分と同じ心理状態にあるだどとどうしてわかるのだろうか。この他人の心の問題「他我(たが)問題」は数百年前から哲学者の間で議論されてきたがほとんど進歩はなかった。しかし現在では，他我問題を研究するための実証的な科学の基盤が整ってきている。

現在では，我々の脳は，他者の心をニューロン単位できめ細かく映し出していると考えられるようになってきた。すなわち他者の心を「ミラーリング」できるらしい。

また，別の観点からこの問題を考えてみると，自閉症のような社会性が欠如してしまう病気もミラーニューロンの機能不全によるものではないかと考える研究者もいる。

ニューロエシックス（Neuroethics 神経倫理学），ニューロマーケティング (Neuromarketing 神経マーケティング), ニューロポリティクス (Neuropolitics 神経政治学) のような新しい分野の研究にもミラーニューロンの機能がかかわってくると考えられている。

### 1.1　ミラーニューロンの発見

サルの前頭葉に F5 という領域がある (図 1)。腹側の運動前野の一部でる。Rizzolatti らは、この領域から奇妙なニューロン活動を記録した。このニューロンは、サルが運動するときに活動するだけでなく、サルがヒトが行う同じ運動を見ているときにも活動した。図 2,3 に彼らが記録したニューロン活動の一例を示す。図中のニューロンはサルが指で餌を摘むときに活動している。図 2 の右側、ヒトがトレーを持ってサルの手がトレー中央にのびている絵の下がそのときのニューロン活動である。このニューロンはヒトが指で餌を摘むときにも活動した。図 2 の中央付近、ヒトの左手がトレーを持ち右手で餌を摘んでる絵の下がそのときの活動である。

図 3 はの中央付近は図 2 と同じようにヒトの左手がトレーを持っているが、右手はペンチを持ち、そのペンチで餌を摘んでいる。この状況下ではこのニューロンは活動しなかった。このテスト期間中もサルが自分で餌を摘むときはニューロンは活動していた。図 3 の右側はそのときの様子を図示したもので、その下はそのときのニューロン活動である。更に図 4 では暗闇でサルが餌をつまんでも、このニューロンは活動した。

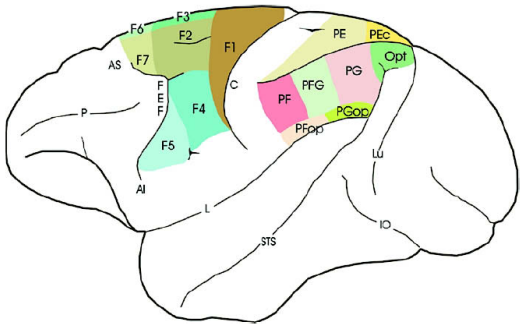


図 1: サルの F5 領域. 運動前野の一部で腹側前方に位置する。C は中心溝、P は主溝、AS は弓状溝の上枝、Al は弓状溝の下枝。(Rizzolatti et al.,2004 を改変)

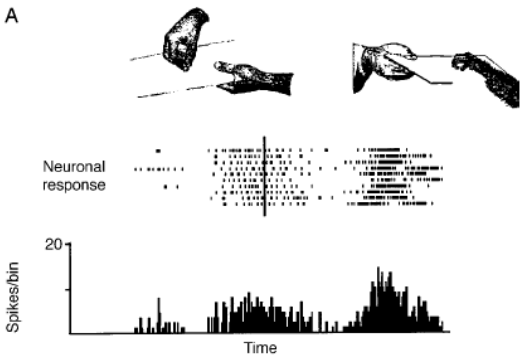


図 2: F5 から記録されたミラーニューロンの例。一番上はヒトとサルの手の動作の絵。それぞれの絵の下がそのときのニューロン活動。中央の絵の下がヒトの動作をサルが見ているときの活動。右の絵の下がサルが自分で餌を摘むときの活動。ニューロン活動の図の上はニューロンの活動電位を点で表してあり、10 回同じ動作を繰り返したときの活動である。下の図は 10 回繰り返したときの活動を加算平均したヒストグラム。(Rizzolatti et al.,1996 を改変)

これらのデータから、このニューロンが餌を摘むとき、視覚像がなくても自分が餌を摘む運動をするときには活動することがわかる。別な言い方をすれば運動情報を扱っているニューロンである。一方、図２と図３から、ヒトが餌を指で摘む動作をしているのを視覚的に見ているときに活動している。しかし、ヒトの動作を見ているときの活動は、視覚像そのものに活動しているのではない。それは図４で視覚像なしに活動したことから得られる推論である。ではヒトの動作を見たときに何故活動したのであろうか。これは、ヒトが行った動作が自分が行った動作と同じであることを理解していたと解釈できる。このような活動を示すニューロンをミラーニューロンという。

現在では，F5 領域の約 20%がミラーニューロンであり，80%は違うことが分かってきている。ミラーニューロンの重要性は，ニューロンが知覚，認知，行動といった枠組みを単一ニューロンが越えることができるという方法論上の大転換を意味しているからである。

シミュレーション説という考え方がある。他人が何をしているのかを理解するためには，自分もその行動をシミュレートしていなければならないというものである。例えば，その人が恋をしていた場合には，自分も恋をしているかのように振る舞わなければならないというものである。すなわち，他人の心を理解するために必要なシミュレーション過程を支える神経基盤

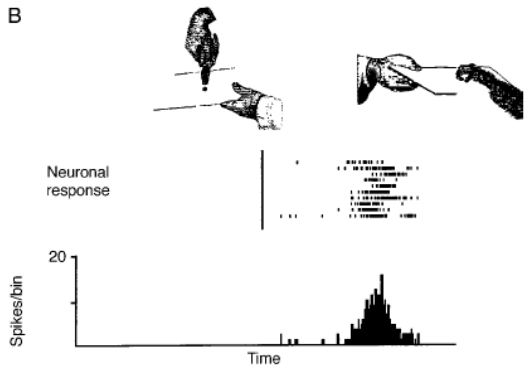


図 3: このニューロンはヒトが道具で餌を摘むときには活動しなかった。(Rizzolatti et al.,2004 を改変)

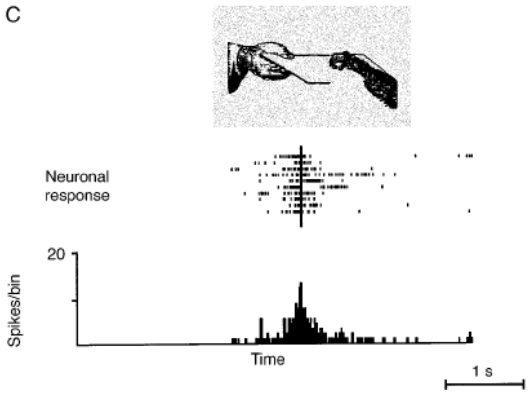


図 4: このニューロンは暗闇でサルが餌を摘むときにも活動した。(Rizzolatti et al.,2004 を改変)

がミラーニューロンなのではないかという考え方がある。

## 1.2 キャノニカルニューロン

キャノニカル（正準）ニューロンというニューロンも発見されている。これは，サルがものをつかむという行動（把持行動）をするときに発火するニューロンであると同時に，つかむことができる対象を見たときでも発火する。キャノニカルニューロンにせよミラーニューロンにせよ,「行動と知覚とは完全に独立した過程であり，それぞれが格納されている脳内領域は異なっている」という従来の考え方とは矛盾するものである。

サルにせよヒトにせよ，他者がリンゴを拾い上げるのを見たときには必ず，脳内で自分がリンゴをつかむのに必要な運動計画を開始させている（ミラーニューロンの活性化)。同様にして，ただのリンゴを見た時でも，リンゴをつかむのに必要な運動計画を開始させるものと考えられる（キャノニカルニューロンの活性化)。リンゴを手に取って食べるために必要な把持行動と運動計画とは，リンゴに対する我々の「理解」そのものに本質的に結びついている。ミラーニューロンとキャノニカルニューロンの発火パターンは，知覚と行動とが脳内で分離していないことを示している点が重要である。

## 1.3　言語野との関係

ミラーニューロンが記録されたサルの F5 は、ヒトの運動性言語野（ブローカ野）に相当する位置にあることから、ヒトの言語機能との関係がいろいろ議論されている。可能な解釈のひとつは、ミラーニューロンは、相手の脳が行っている運動制御の内的な状態を推定し、自分の運動の表象を使ってリハーサルする役割を持っているというものである。これは模倣にもつながる機能である。模倣は乳幼児が言葉を覚えるときの初歩であることから、ミラーニューロンは非言語コミュニケーションの基盤にもなると見られている。しかし、これらの考察はあくまでもサルのデータからの推論の結果であり、ヒトの言語機能で同様のニューロン活動が機能しているという証拠は今のところない。だが、サルの F5 に対応する人間の大脳領野は Brodmann の 44 野 45 野すなわちブローカの言語野である。

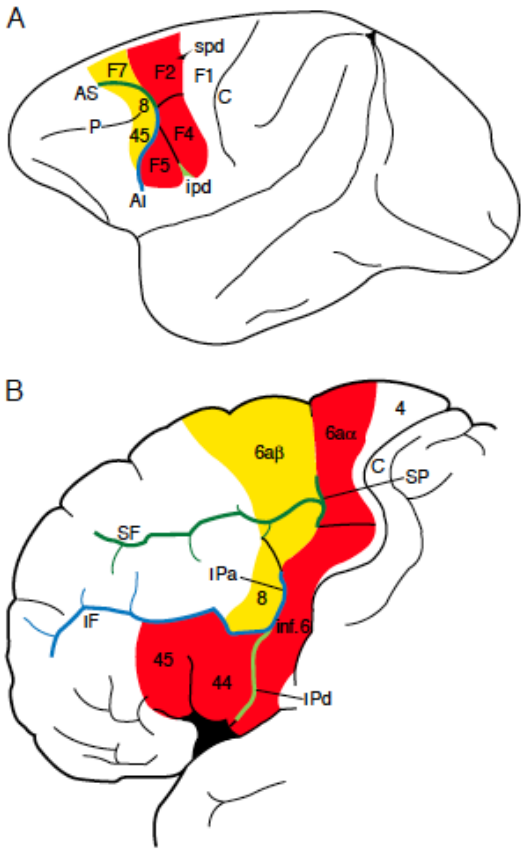


図 5: サルの脳とヒトの脳との対応関係（Rizzolatti et al.,2004 を改変）

ミラーニューロンの特徴を持つニューロン活動は運動前野の F5 領域の他、頭頂連合野の 7b 野の一部である PF 野や側頭連合野の上側頭溝領域前方（STSa）でも記録されている。これらの領域が構成するネットワークが、この特徴ある活動をつくりだしていると考えられている。さらに、模倣や自己と他者の区別などの課題遂行中のヒトの脳機能イメージングの研究によって、サルと同様の回路が働いていることを示すデータが発表されている。

# 2　意味記憶の構造試論

## 2.1　アルツハイマー症の物体呼称課題における成績低下

アルツハイマー症の初期症状の一つとして、物体呼称課題における障害が挙げられる。患者は椅子のことを尋ねられてテーブルと答えたり、梨のことを果物と言ったりする。このことは意味記憶の構造を考える上で興味深い。同一カテゴリーに属する別のメンバーである椅子とテーブルとを間違えることと、具体的な事物を表す梨をより上位概念である果物と答えることが起こっている。アルツハイマー症の意味記憶構造にはどのような障害を考えればよいのだろうか。

アルツハイマー症の患者の物体呼称課題における障害には、上記の意味記憶の障害説以外にも、二つの別の説明が存在する。一つめは、視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた課題で、線画を用いた方が呼称成績が悪いのである。すなわち、視覚性の障害だと解釈できる。二つめは語彙性の障害の可能性である。高頻度語の想起の方が低頻度語の想起よりも成績がよいという頻度効果が存在する。

アルツハイマー症患者の物体呼称課題における成績低下は、意味性、視覚性、語彙性のいずれの障害によって引き起こされるのであろうか。個々の患者ごとに障害の場所が異なると言う説明の仕方ももちろん可能ではある。しかし、ティペット (Tippett) とファラー (Farah) (Tippett & Farah, 1998) によるニューラルネットワークを用いた研究によれば、意味記憶の障害のみによってアルツハイマー症の物体呼称課題における成績低下を説明できる可能性がある。ニューラルネットワークの特徴は知識の分散表現と相互作用であり、意味的な知識表象と視覚的および語彙的な知識表象は密接に関連しあっているのである。その結果、一つの構成要素、例えば意味記憶に障害があると、その障害の影響は、視覚性の知識にも語彙の知識にも影響を与える可能性がある。

## 2.2　モデル

彼女らのモデルを図 6 に示す。3 つのユニット群、意味、名前、視覚がある。実際には各層の間に中間層が存在するのだが、ここでは中間層の存在は本質的ではないので省略した。例

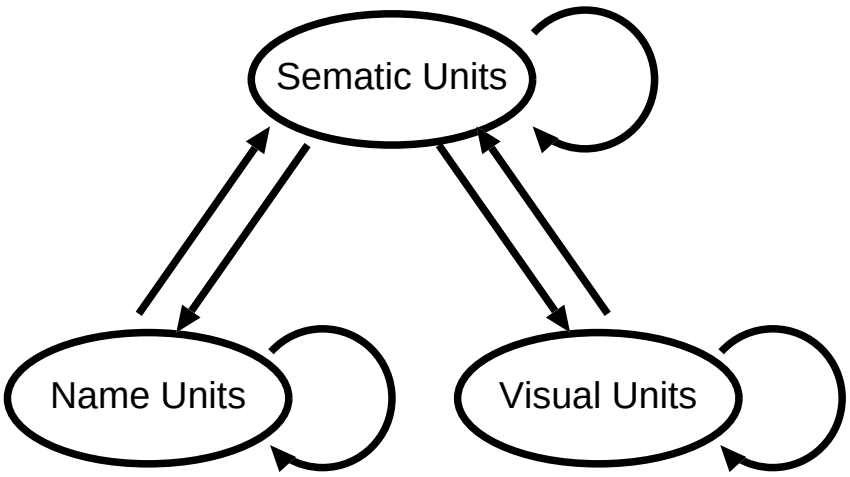


図 6: ティペットとファラーの用いたモデルの概略図

えば視覚ユニット層に入力が与えられると、各ユニットの活性化 (あるいは負の結合であれば抑制) は意味ユニット層に伝播する。意味ユニット層の活性化が、さらに視覚ユニット層と名前ユニット層の活性化に影響を与える。図中の矢印で示されているとおり各ユニット間の結合は層内、層間で双方向である。従って、あるユニットの活性は別のユニットの活性を引き起こし、全体として複雑な活性化パターンを示す。彼女らは視覚ユニット層 (あるいは名前ユニット層) に入力を与えたときに、意味ユニット層と名前ユニット層 (あるいは視覚ユニット層) に、特定の活性化パターンが示されるようにニューラルネットワークを訓練した。

### 2.3 視覚性障害仮説についてのシミュレーション

訓練後、意味層のユニットを除去することで脳損傷がシミュレートされた。脳損傷を受けていないネットワークと脳損傷を受けたネットワークとに対して、視覚刺激が完全である場合と、不完全な場合とで名前ユニット層に現われる活性化パターンがどのように変化するのかが観察された。

結果は交互作用が観察された。正常なネットワークに、不完全な視覚刺激を与えても、名前ユニットに現われる活性化パターンは学習したパターンに近いものが観察されたが、脳損傷を受けたネットワークでは物体呼称成績が著しく障害されたのである。

### 2.4 語彙の頻度効果に対するシミュレーション

続いて、語彙の頻度効果を調べるために、高頻度語と低頻度語との違いをニューラルネットワークの訓練回数の違いとして表現した。視覚性障害仮説についてのシミュレーションと同じように、訓練後、意味層のユニットを除去することで脳損傷が表現された。脳損傷を受けたネットワークの名前ユニット層に低頻度語を提示したときの成績は著しく低下するが、高頻度語を提示したときの成績はそれほど低下しなかった。すなわち、語彙性の障害と考えられてきたような症状を意味層の障害によって再現できたのである。

以上見てきたように、ニューラルネットワークを用いると、意味記憶の障害だけを仮定すれば、視覚性の誤りも語彙性の誤りも説明できる。すなわち従来から考えられてきた課題成績によるアルツハイマー症の障害の分類に全く新しい視点が与えられるのである。それでは、意味記憶そのものの構造はどのようになっているのであろうか。

## 3 意味記憶の構成 —生物、非生物の二重乖離—

認知心理学でしばしば話題になる記憶表象論争に関して、意味記憶は、個々の対象についてカテゴリーごとに構成されているのかそれとも、それともモダリティーごとに構成されているのか、という論争がある。ファラー (Farah) とマクレランド (McClelland)(Farah & McClelland, 1991) が行なったニューラルネットワークによる研究によれば、モダリティーに依存した意味記憶表象を考えれば、カテゴリーに基づく意味記憶表象は説明できる。

### 3.1 神経心理学的症状

実際の脳損傷患者の中には、動物や植物などの生物の知識について障害がある一方で、非生物の知識については健常のまま保たれている患者が存在する (Warrington & Shallice, 1984)。古典的な二重乖離の原則から、生物と非生物の知識の脳内での意味記憶には、生物と非生物と

を独立に表象している意味記憶が存在すると仮定される。しかし、ウォリントン (Warrington) とシャリス (Shallice) は、生物の知識と非生物の知識との間で選択的な障害が起こるのは、異なる感覚運動経路からの情報の重みの差異を反映しているためではないか、と述べている。すなわち、生物は主に感覚的な性質によって互いを区別することが多いが、非生物は主に機能によって分類される。ある動物、例えばヒョウは、他の肉食動物と比べて主に視覚的な特徴によって差別化される。これとは対照的に、机の知識については、他の家具との違いを記述するときには主に機能、すなわち何のために使うのか、によって差別化される。それゆえ、障害のある知識と健全に保たれている知識との違いは、生物–非生物の違いなのではなく、対象を記述している特徴が感覚–機能の違いであるかも知れない。

ファラーとマクレランド (Farah & McClelland, 1991) のモデルは上記の感覚–機能仮説が意味記憶障害を説明できることを例示するために作成された。

### 3.2 モデル

彼女らのモデルを図 7 に示す。3 つのユニット群、記憶を表現する意味記憶系と、入出力を表現する二つの周辺系、視覚ユニット群と言語ユニット群とがある。言語ユニット群と視覚ユニット群との間を除いて、全てのユニットに群間および群内結合が存在した。

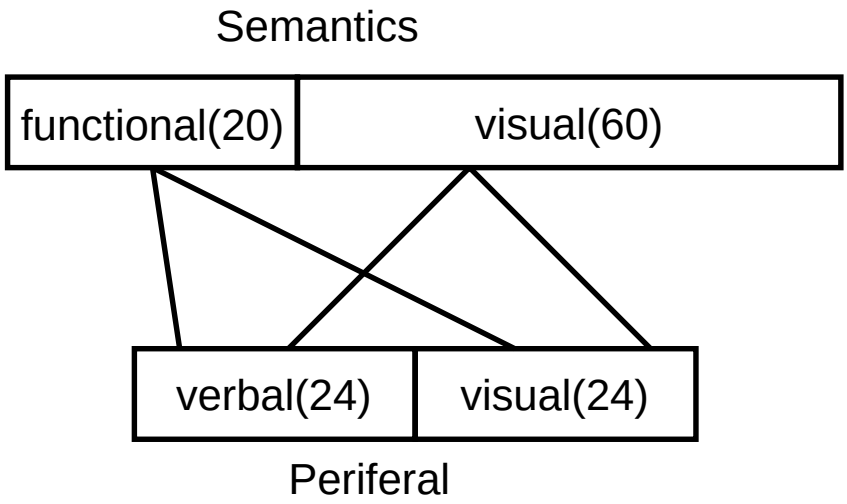


図 7: ファラーとマクレランド (1991) の意味記憶モデルの概念図。カッコ内の数字は数値実験で用いられたユニット数を表す。意味記憶内で機能的記憶と視覚的記憶のユニット数が異なるのは、彼らの論文中の実験 1 (心理実験) の結果を反映している。

このモデルに生物と非生物を表す刺激が提示された。生物と非生物とを表す項目は、視覚情報と機能情報との比率が変えられた。生物項目では平均して 16.1 の視覚意味記憶ユニット、2.1 の機能意味記憶ユニット。非生物では 9.4 の視覚意味記憶ユニット、6.7 の機能意味記憶ユニットを用いて表現された。視覚パターンが提示されたときには対応する意味記憶パターンと言語パターンが産出されるように、また、言語パターンが提示されたときには対応する意味記憶パターンと視覚パターンが産出されるよに訓練された。各訓練試行では、生物、もしくは非生物に対応する視覚入力や言語入力が言語ユニット群あるいは視覚ユニット群に対して提示され、ネットワークは解が安定するまで活性値の更新が行なわれた。

### 3.3 破壊実験

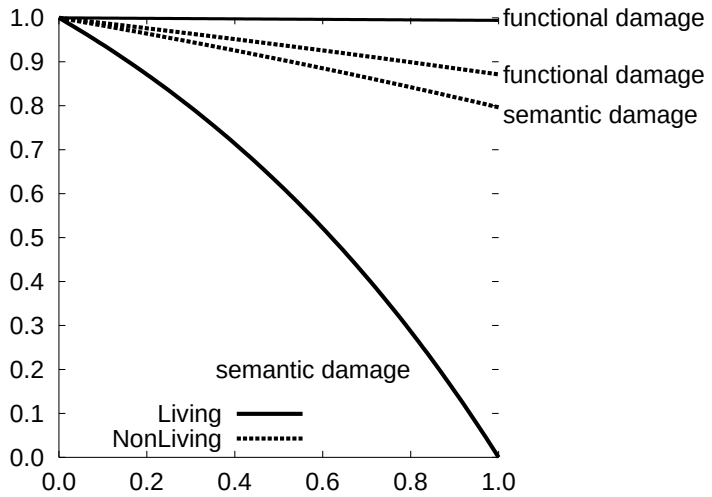


図 8: 生物–非生物別の意味記憶内の損傷の程度と課題成績との関係 (ファラーとマクレランド (1991) の表 3 と図 2 より改変)。彼らは各条件で 0, 0.2, 0.4, 0.6, 0.8, 0.99 の 6 点について各５回づつしか数値実験を行なっていないため実際の曲線は滑らかにならない。そこで、指数関数に回帰させてプロットしなおした。指数回帰を用いれば唯一のパラメータを変化させることで４つの条件に対応する曲線を描くことができる。

彼女らは、モダリティーに依存した意味記憶障害、すなわち、耳で聞いたときには理解できるが、目で見たときには特定のカテゴリーについての知識に障害を生じる患者のシミュレーションを行なった。さらに、ネットワークに雑音を加える方法によって脳損傷を表現し、シミュレーションを行なった。彼女らのシミュレーション結果は、モダリティー依存の意味記憶構造を用いれば、カテゴリー依存の障害を説明できることを示している。すなわち生物、非生物という異なるカテゴリーに属する事物は脳内で異なって表象されているのではなく、意味記憶の構造はは入力モダリティーに依存して形成されているという結果が得られたのである。

## 4 ニューラルネットワークによる障害の再解釈

脳損傷を扱った認知障害の研究では二種類の実験を使って脳損傷患者の障害の場所を特定することが試みられてきた。一つめは患者の示す誤りの種類を分析することである。例えば、物体の呼称課題において、視覚的に似ているものを答える誤りについては視覚性の誤りと分類し患者の視覚機能に問題があると推論する。同様に意味性の誤りについては意味記憶に障害があると判断される。二つめは課題の難易度を操作して障害の場所を特定することである。例えば呼称課題において、低頻度語の呼称成績が選択的に障害されていれば、語彙に障害があるとする。視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた物体呼称課題で、線画を用いた方が呼称成績が悪ければ、視覚性の障害とみなす、などである。

ところが、今回説明したようにニューラルネットワークモデルによる研究では、認知機能の局在を示す脳損傷患者のデータと、その認知機能を推論する伝統的な認知神経心理学的手法に疑問を投げかけている。ニューラルネットワーク研究によって、従来からの神経心理学的障害分類論に本質的な変更が迫られているように思えるのだ。

# 5　自己組織化の意味と意味の自己組織化

「自己組織化」とは非常に壮大なテーマである。この問題に直接答えるのには私には荷が重すぎるし、この連載の主旨からもはずれてしまう。だがあえて自己組織化の定義を試みれば「自己組織化とは、経験と環境の関数として基本構造が変化し、合目的的システムができあがること」と定義することができるだろう。

例えば、人間は自己組織化システムである。だれもが一個の有精卵から次第に複雑な構造を発生させて行ったのだから。もっとも、すべての生物は自己組織化システムであるし、太陽系も自己組織化システムだと言うことができるかも知れない。さらに初めの初めから始めるとすれば、30 億年前原始地球の原始スープの中から長い年月をかけて自己複製を始めた生物の発生にさかのぼることができる。原始生命の出現に超越的な創造者の存在を仮定するべきなのだろうか？それとも、現在の生物の持つ自己複製機能の創発を認めるべきなのだろうか？現代生化学の研究成果は、超越的な創造者の存在を仮定しない生命発生のシナリオを描き始めているように思われる。単純な自己増殖機能を持ったタンパクから、やがて細砲が作られ単細胞生物へ、さらに多細胞生物へ、さらに陸上へと進出し、火を発見し、文字を発明し、知的活動を行なうようになった実例が今の私たちである。生物が自身の知的活動をシミュレートするようになるまでには、多様なレベルでの自己組織化が行なわれて来たのだと想像できる。

現代的な意味でのニューラルネットワークにおいては、上記のような意味での「自己組織化」は実現されていない。現在のニューラルネットワークにできることは、極論すれば、外界の統計的構造を獲得することができるという点である。もうすこし具体的にいえば、外部入力の統計的構造を内部のシナプス伝導効率の変化として表現することができる、ということである。ここから、知的な活動を創発できることまでの間には厖大な距離がある。だが、外界の情報すなわちデータの相互関係を効率良く表現することは情報科学の分野でも中心的な問題であり、おそらくこのような能力が脳の働きの特徴の１つであるということができる。今回は自己組織化という壮大なテーマの入口、外界の情報から意味のある構造を作りだす、という点に的を絞って説明しよう。

## 5.1　トポグラフィックマッピング

外界の構造が脳内の地図として表現されていることは一般に知られた事実である。例えば、網膜と第一次視覚野の間には連続的な１対１対応が存在する。鼓膜の周波数選択特性と第一次聴覚野との間にも対応関係が見られる。同様に体表面の感覚と体制感覚野の間にも対応関係が見られる。すなわち感覚器官と第一次感覚野との間の神経結合は、類似した刺激に対して皮質上の同じような位置に対応する受容野を持つことが知られている。このような 2 つの神経場間の連続的な結合関係のことをトポグラフィックマッピング topographic mapping と言う。このような構造は、大まかには遺伝子によって決定されているが、細かい構造については神経回路の自己組織化によって達成されると考えられている。

## 5.2　自己組織化の意味するもの

大脳皮質全体の 10 % を占める第一次感覚野で起こっていることの類推から、特定のカテゴリーにおける知識表現が脳の各部位の位置関係として表現されているという可能性があるだろうと考える。すなわち、さまざまなレベルの情報表現の自己組織化に対して、たった 1 つの同じ機能的原理が働いているのではないか、という仮説が提起できる。第一次感覚野で表現されている情報表現と同じ機能的原理が、知的なレベル (各種の連合野、あるいは前頭葉) で

も同じであると考えてはいけない理由はないはずである。仮に、この同一の機能的原理が高次の知的活動のためにも働いているのなら、低次の感覚受容野から、階層的に高次の連合野にいたるまで自己組織化によって我々の知的活動のある部分が説明可能なのかも知れない。自己組織化によって高度に抽象的な概念が階層的に重ね合わさっていた場合にどのようなことが起こるのだろうか。第１次感覚野が物理的な特徴量を表現し、第２次感覚野が具体的な概念を表現しているとしたら、連合野は抽象的な概念を表象しているのかも知れない。連合野の連合野である前頭葉では概念の概念の概念が形成されているというのは誇張のしすぎなのだろうか。概念の概念の概念は知的な能力とみなしても良いと思う。すなわち自己組織化が多段階に重なることによって抽象度の高い知的能力が創発すると考えても良いのではないだろうか。

## 5.3 意味の抽出

ランダウアー (Landauer) とデュマス (Dumais)(Landauer & Dumais, 1997) は、百科事典のすべての文章における単語の見出し語項目との間の共起関係に特異値分解を適用し、数百個の次元からなるベクトル表現を構成した。このベクトル表現によって単語間の類似度を定義し、TOEFL の類義語問題に結果を適用することで人間の受験者に近い正答率が得られることを示している。彼らによればベクトルの次元数を 300 としたとき一致率が最大になるという。このことから人間の意味処理として 300 次元程度の意味空間を用いることでコンピュータに人間に近い振る舞いをさせることができるという結果が得られている。このことは従来曖昧な定義であった意味に対して計量的なアプローチが可能であることを示していて興味深い。

## 5.4 自己組織化アルゴリズム

ランダウアーとデュマスの使った特異値分解とは、数学的手段であり、固有値問題と関連が深い。固有値問題は、多次元の情報を情報の損失を最小にしながら低次元の情報に変換する情報圧縮のために使われたりもする。従って与えられたデータの固有値問題の解を自己組織的に学習して解くニューラルネットワークがあれば、ランダウアーとデュマスたちの示した結果をニューラルネットワークでも表現できることになる。固有値問題を解く自己組織化ニューラルネットワークには、ヘッブ (Hebb) の学習則、およびヘッブの学習則を拡張したオヤ (Oja) の学習則 (Oja, 1988)、オヤの学習則を拡張したザンガー (Sanger) の学習則 (Sanger, 1989) などが知られている。固有値問題とはどのようなものかを説明せずに、この文章を読んでも意味不明であるとは思うが、固有値問題とは情報の圧縮であり、抽象化である、と思っていただいてよい。ヘッブの学習則を使うとシナプスの結合係数が最大固有値に対応する固有値ベクトルの方向と一致し、オヤの学習則を使うとその固有ベクトルが１に規格化され、ザンガーの学習則を使うと望む数だけ固有ベクトルが大きい順にとりだせるということである。数式を用いずにこれらのことを説明するのは大変なのだが、多数のニューロンと結合を持つ一つのニューロンを考えたとき、このニューロンへのシナプス結合係数の変化は、このニューロンの発火率とこのニューロンへ信号を送っているニューロンの発火率の積で表されるというのがヘッブの学習則であり、ヘッブの学習則に正則化項を取り入れたものがオヤの学習則であり、オヤの学習則を多層化したものがザンガーの学習則なのである。

数式を使わないで説明したため、いささか面妖な日本語になったがお許し願いたい。要するに初めに戻って、外界の統計情報を効率良く学習するニューラルネットワークモデルが実在するのだと言いたいのだ。そしてランダウアーとデュマスの結果を信じれば 300 次元程の意味次元を考えれば人間の知識を表現することができ、また、それは自己組織化アルゴリズムを用いて実現可能だと言うことである。

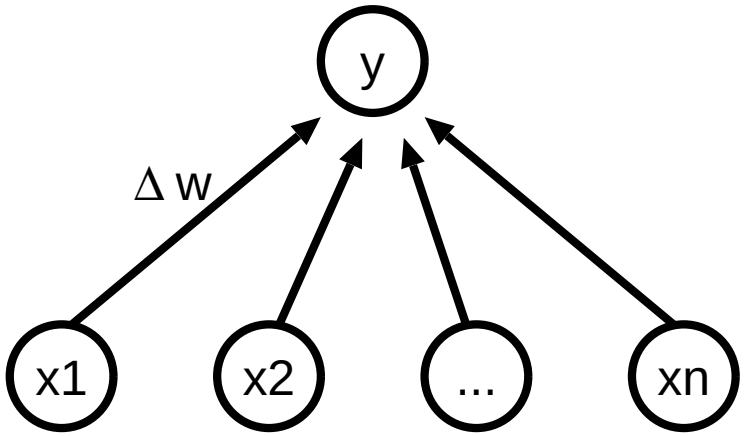


図 9: 自己組織化の例。2 層のネットワークを考え、下位層の各ユニットと上位層にあるユニットとの結合係数 $w$ が固有ベクトルに対応するような自己組織化アルゴリズムが提案されている。$\Delta w_i = \eta y x_i$ がヘッブの学習則であり、$\Delta w_i = \eta y(1-y)x_i$ がオヤの学習則であり、$\Delta w_i = y(x_j - \sum x_k w_k)x_i$ とするのがザンガーの学習則である。

## 6 NMF による意味の分解

同じような発想から非負行列因子化 (NMF) と呼ばれる手法も最近注目を集めている (Lee & Seung, 1999)。NMF は入力データを構成する基底を抽出する自己組織化アルゴリズムである。実際に NMF を顔画像処に用いていろいろな人物の正面顔を入力した場合では、顔を構成するパーツ、目や鼻や口のような画像が基底として抽出された。NMF は基底と展開係数の成分が共に非負であるという性質を持っている。NMF は基底と展開係数を更新することによって外界情報の持つ性質を抽出する自己組織化アルゴリズムの一手法である。NMF を事典の各項目に対して応用した例では、例えば「アメリカ合州国憲法」は大統領、議会、権力などの各因子を展開係数を用いて加算した形で表される。このように各概念が、下位概念 (因子) とその重みである展開係数の積とで表現される点が NMF の特徴である。

NMF の応用を示した簡単な例が `http://www.twcu.ac.jp/~asakawa/nmf/`にあるので参照して頂きたい。ここでは小学生が学習する学習漢字 1006 字に対して NMF を実行し、ごんべんやしんにょうなどのパーツが基底として抽出されたことが示されている。

NMF を使えば、ランダウアーとデュマスの使った語彙も、各々の単語の下位のパーツとなる語彙とその展開係数で表すことが可能であろう。「意味の自己組織化」の試みは確実に進歩してきているといってよいだろう。これは一昔前にくらべてコンピュータの処理速度と記憶容量とが十分になってきていることと関係している。いよいよ面白い時代になってきたと言えるのではないだろうか。

## 7 カテゴリー特異的な意味記憶の障害を階層構造と自己組織化マッピングを用いたモデルで説明する試み

物品や対象の認識や呼称に選択的な障害を持つ脳損傷患者は，これら概念の認識過程にかかわる重要な手がかりを与えてくれる．とりわけある特定の対象の認識や呼称についてだけ障害のある「意味記憶のカテゴリー特異性」障害は示唆に富む．生物，果物野菜，道具などが選

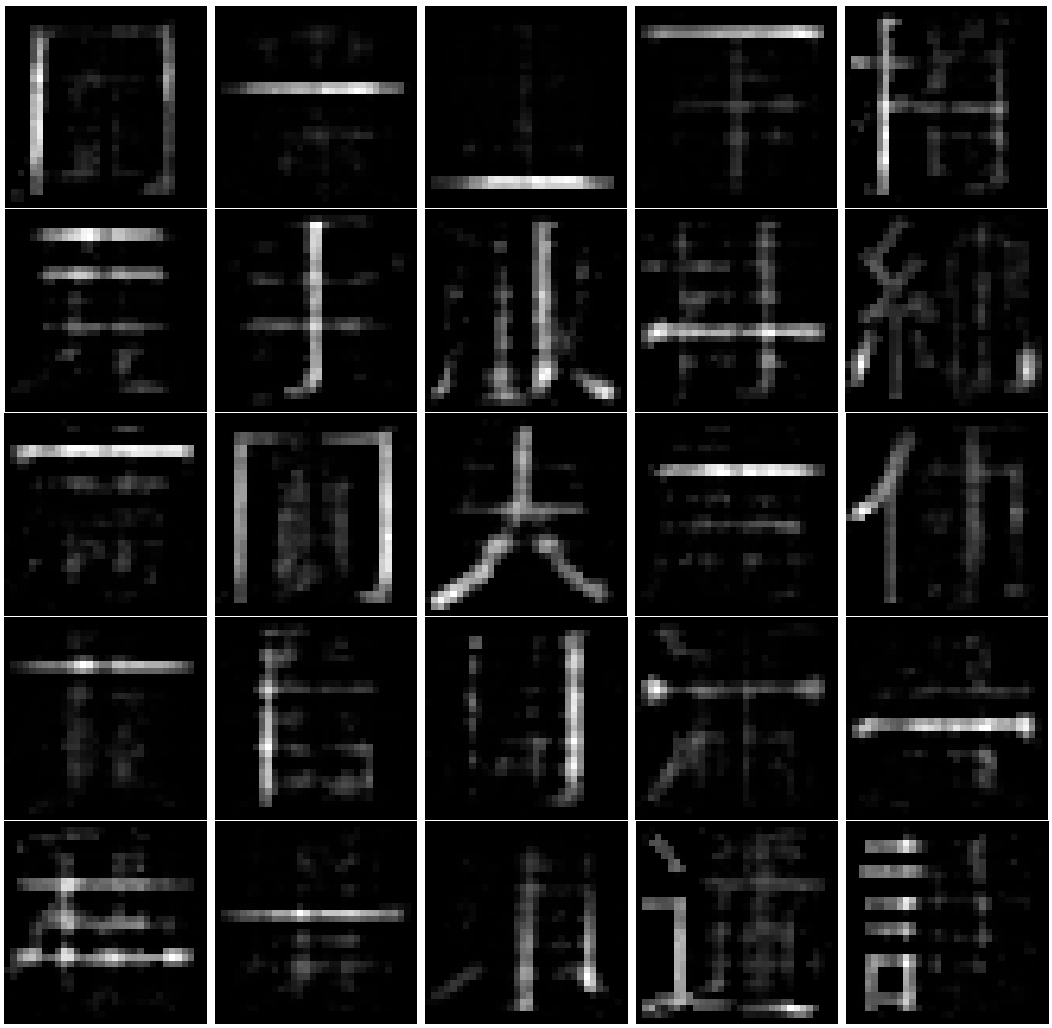

図 10: NMF で学習漢字 1006 文字を処理した結果

択的に障害されることが知られている．脳損傷後に現われるカテゴリー特異性を説明するためにはどのようなモデルが必要なのだろうか？

本稿では，これまで提案されて来た仮説が，２層の自己組織化マップ (SOM) とその間を繋ぐ Hebb 則による連合によって表現できるという単純なモデルを提案し，簡単な数値実験を行なった．

現在までに提案されて来た意味記憶に関するカテゴリー特異性を説明するモデルは３つに大別される．(1) Warrington & Shallice (1984) は、カテゴリー特異的な障害は対象の認識のためには種類の異なる情報が必要であると考えた．生物を区別するためには感覚的情報が決定的に重要であり，一方，機能的情報 (例えば，どうやってそれを使うのかといった) が非生物を区別するためには重要であると考えた．この「感覚/機能」仮説は，近年におけるカテゴリー特異的な障害の説明のためにもっとも広く用いられて来た仮説である．(2) Santos & Caramazza (2002) は，概念が「領域特殊性」によって体制化されているが故に概念毎に選択的な障害が発生すると論じている．(3) Humphreys & Forde (2001) たちの Hierarchal Interactive Theory (HIT) 仮説は，症例のパターンを説明する際に，関連したカテゴリーの事例の類似性が関与していることが強調されている．彼らは生物についてのカテゴリー選択的な障害は，これらの事

例が知覚的に類似した事例を持つカテゴリーに属しているからだとする．

感覚/機能仮説に基づくにせよ，領域特殊性仮説に基づくにせよ，類似性仮説に基づくにせよ，ある特徴ベクトルによって表現された対象が互いに近傍に表象され，その損傷によってカテゴリー特異的な障害が発生すると考えれば，自己組織化写像による入力ベクトルの体制化機構を想定することで，これまでに提案されたモデルの要求を満足すると考えることができよう．なぜなら，感覚情報によって表現された概念と機能情報によって表現された概念とが異なる位置に表象されることは「感覚/機能」仮説を実現したものであり，結果として形成される 2 次元布置がカテゴリー毎に体制化されているとすれば「領域固有性仮説」を実現したことになり，同時に類似性の判断に基づく類似性仮説を満たしていることになるからである．

本稿では Hinton & Shallice (1991) で用いられた文字の視覚情報と概念とを結びつけ深層失読をシミュレートするために用いられたデータを用いてモデルの検証を行った．データは室内の物品，動物，身体の部位，食物，室外の対象の 5 つのカテゴリー，計 40 事例である．各事例の意味ベクトルは例えば max-size-less-foot や has-legs などの項目からなる 68 次元の 0,1 ベクトルであり，語彙ベクトルは 3 文字あるいは 4 文字からなる単語の 0,1 表現からなる 28 次元ベクトルであった．意味，語彙それぞれに対して SOM を適用し，入力ベクトルの 2 次元の布置を学習させ，同時に SOM の winner–take–all 回路によって勝者となった 意味 SOM ユニットと語彙 SOM ユニットの間で Hebb 則による学習を行わせた．

本稿で用いたアルゴリズムは以下の通りである．各入力ベクトル $\boldsymbol{x}$ に対して最もマッチした SOM のユニット $i$ を winner–take–all 回路

$$\boldsymbol{x} \mapsto c\left(\boldsymbol{x}\right) = \underset{i \in \{1,\ldots,N\}}{\operatorname{argmin}}\ d_w\left(\boldsymbol{w}_i, \boldsymbol{x}\right). \tag{1}$$

を定義し，学習則

$$\Delta \boldsymbol{w}_i(t) = -\alpha(t) n\left(i, c\left(\boldsymbol{x}\right)\right) \frac{\partial d_w\left(\boldsymbol{w}_i, \boldsymbol{x}\right)}{\partial \boldsymbol{w}_i} \tag{2}$$

を適用する．ここで $n(x, y)$ は近傍関数であり $n(x, y) = \exp(-|x - y|^2/\sigma^2)$ とした．$\alpha(t)$ が学習係数であり，学習回数 $t$ の単調減少関数である．ここでは $\alpha(t) = \eta\left(1 - t/t_{\mathrm{tmax}}\right)$ とした．$\eta$ は定数である．$\alpha$ の存在によって SOM の収束が保証される (Kohonen, 1997).

2 つの SOM の勝者ユニット間を結ぶ Hebb 則においては活性値を $[0, 1]$ に制限するために入力 $x$ に対する応答関数として $y = 1 - \exp(-\beta x)$ を用いた．従って通常の Hebb 則にではなく，2 つの SOM の勝者同士を結ぶ学習の際には

$$\delta w = \beta \exp(-\beta x) \tag{3}$$

という学習則を用い，それ以外のときには指数関数に従って忘却させることにした．

$$\delta w = \gamma w \tag{4}$$

忘却に指数関数を用いたのは，SOM の学習途中に勝者が変化した場合，過去の情報を速やかに消去し，新たな結合を学習しなければ意味 SOM と語彙 SOM との間に一対一対応の学習が成立しにくくなるからである．$\beta$ が大きければ新たに勝者となった SOM のユニット間の学習が速やかに成立し、$\gamma$ が大きければ忘却しにくくなることを意味する．これは意味 SOM と語彙 SOM の学習時に勝者ユニットが一貫している場合には問題を生じないが，SOM の学習にともなって勝者ユニットに移動や変化が生じた場合，意味と語彙との間に混乱が生じることを意味する．

20 行 20 列の SOM に Hinton & Shallice (1991) の 5 つのカテゴリー 40 事例を学習させた意味 SOM の結果を図 11 に示した．図から 5 つのカテゴリーが分離されて表象されている

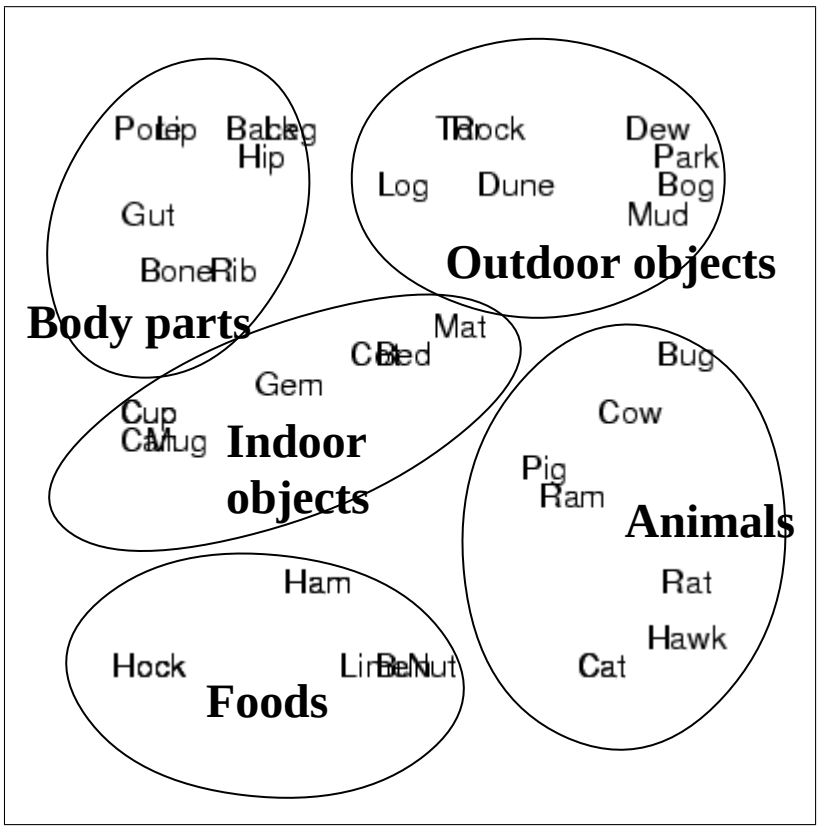


図 11: 意味 SOM の結果

ことが分かる．このことから SOM がカテゴリー特異性を説明するモデルとして有効であることを示唆していると考えられる．すなわち，図 11 において動物を表す領域に損傷が生じると動物概念に特異的な意味記憶の障害が生じるものとみなすことができよう．

一方，Hebb 則による意味 SOM と語彙 SOM との連合の Hinton グラフの一例を図 12 に示した．この結果は，2 つの SOM の学習と提案した Hebb 則の変形による学習則によって意味と語彙との間に一対一連合が成立しうることを示したものであり，本稿で提案したモデルの妥当性を示すものと言える．しかし，図 12 右からもわかるとおり，完全に一対一対応が取れる場合でばかりではないことも確かである．これには二通りの解釈が可能である．一つは，学習パラメータ $\beta$, 忘却パラメータ $\gamma$ の決定問題である可能性である．もう一つの可能性は SOM によって意味と語彙とを徐々に学習していく過程で，両 SOM 間の対応問題を同時に学習していくという本モデルが持つ機構そのものに起因する限界である．

上述のようなパラメータの恣意性，もしくは限界が考えられるものの，カテゴリー特異性を説明する試みとしてモデルが実装できた意義は大きいと考える．とりわけ，感覚/機能，領域固有性，そして類似性仮説という 3 つの仮説はニューラルネットワークモデルとして実装した場合，一つのモデルを別の側面から論じているものであると言う可能性を示すものであることは，今まで論じられてこなかった．今後この観点を考慮して検討を進めて行く必要があるだろう．

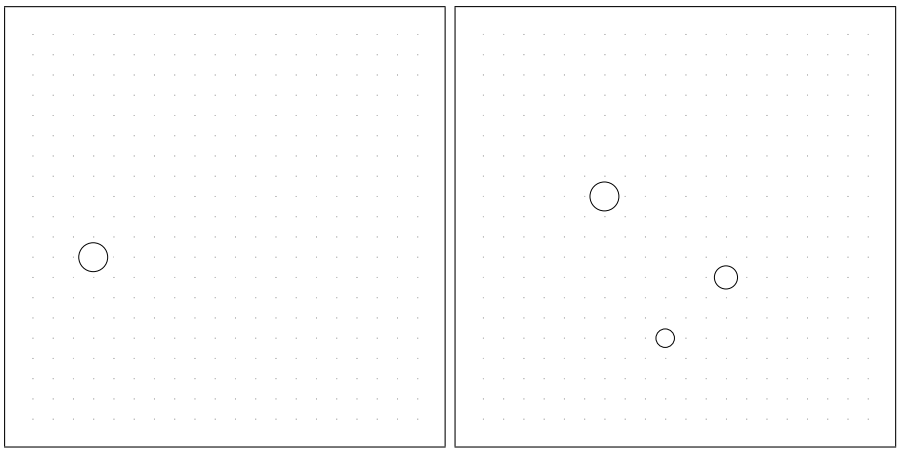

図 12: 2 つの SOM を繋ぐ結合の Hinton グラフ

# References


Chapin, J., Karen, A., Moxon, S., R., Markowitz, & Nicolelis, M. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, *2(7)*, 664–670.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, *91*, 176-180.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339–357.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortyex. *Brain*, *119*, 593-609.

Georgopoulos, A., Schwartz, A., & Kettner, R. (1986). Neuronal population coding of movement direction. *Science*, *26(233)*, 1416–1419.

Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, *98*(1), 74–95.

Humphreys, G. W., & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *Behavioral and brain sciences*, *24*, 453–509.

Iaroboboni, M., Woods, R. P., Barass, M., Bakkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526-2528.

Kohonen, T. (1997). *Self-organiziing maps second edition*. Springer.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquistion, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lebedev, M. A., Nicolelis, M., Beggs, J. M., & Plenz, D. (2003). Brainmachine interfaces: past, present and future. *Trends in neuroscience*, *29*, 536–546.

Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Nicolelis, M., & Chapin, J. (2002). Controlling robots with the mind. *Scientific American*, *287*, 24–31.

Oja, E. (1988). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, *15*, 267-273.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience*, *21*, 188-194.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169-192.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141.

Sanger, T. (1989). Optimal unsupervised learning in a single–layer linear feed–forward neural network. *Neural Networks*, *2*, 459-473.

Santos, L. R., & Caramazza, A. (2002). The domain-specific hypothesis: a developemental and comparative perspective on category-specific deficits. In M. E. Forde & G. W. Humphreys (Eds.), *Category specificity in brain and mind* (p. 1-23). Psychology Press.

Tippett, L. J., & Farah, M. J. (1998). Parallel distributed processing models in alzheimer's disease. In R. W. Parks, D. S. Levine, & D. L. Long (Eds.), *Fundametals of neural network modeling: Neuropsychology and cognitive neuroscience* (chap. 17). MIT press.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairment. *Brain*, *107*, 829–854.

Wessberg, J., Stambaugh, C., Kralik, J., Beck, P., Laubach, M., Chapin, J., et al. (2000). Real–time prediction of hand trajectory bye ensemlbels of cortical neurons in primates. *Nature*, *408*, 361–365.