書記素音韻変換過程に関与する意味系の貢献を評価するためのモデル

浅　川　伸　一

はじめに

本稿では、文字を読む際の脳内過程を考える。文字書記素（orthography）を読み（Phonology）へ変換する計算論的モデルについて考察する。言語は種々の認識活動において重要な役割を果たしている。O'Reilly & Munakata（2000）によれば、脳内における言語活動は知覚野、運動野、および連合野の間の相互作用として実現されており、これらの領野間の相互結合によって多様な言語現象を説明できると考えられる。例えば印刷された文字を見て、その文字の読みを想起したり、その文字の意味を考えたりする場面を考えれば、このことは容易に納得できるだろう。言語を理解する上で、言葉が脳内でどのように表象されているのかという問題は、神経科学、認知科学、心理学、生理学、言語学にとって重要な問題であり、多くの関心を集めてきた研究テーマである。

読みについて、従来提案されてきたモデルとしては、Morton（1980）の logogen model, McClelland & Rumelhart（1981）と Rumelhart & McClelland（1982）による Interactive Activation model, Coltheart, Curtis, Atkins, & Haller（1993）, Coltheart & Rastle（1994）, および Coltheart, Rastle, Perry, Langdon, & Ziegler（2001）による Dual Route Cascaded model, Seidenberg & McClelland（1989）と Plaut, McClelland, Seidenberg, & Patterson（1996）による Triangle model などがある。1990 年代は Coltheart et al.（1993）の記号処理的な二重経路モデル（Dual Route Cascaded model）とニューラルネットワークによって表現された Plaut et al.（1996）によるトライアングルモデル（Triangle model）との間の論争の時期だったといって良い。なおトライアングルモデルという名称は Patterson et al.（1996）などによる。上記の文献に加えて、Plaut, McClelland, & Seidenberg（1995）, Seidenberg, Plaut, Petersen, McClelland, & McRae（1994）, Plaut（1997）などを読めば、この論争の模様が伝わってくる。以降、本稿では Dual Route Cascaded model を二重経路モデル、Triangle model をトライアングルモデルと表記する。しかし、上記のモデルでは意味系の貢献の程度を定量的に評価することができないという問題が指摘できる。そこで、本稿では意味系からの貢献を評価する機構を組み込んだモデルを提案する。

後述するように欧米言語の表記体系と日本語のそれとの相違から、脳の障害に起因する

言語に関する諸症状の発生率には相違には認められる。日本語の表記体系にしても、漢字と仮名との乖離は、２つの表記体系を持つ日本語の特徴である。この日本語の特徴と欧米言語との表記体系の違いを理解するためにも、計算論的モデルによって考察を深めることは、一定の意義があるものと言えよう。

関連する部位およびこれまでのモデル

関連する部位

機能的脳画像研究(たとえば櫻井、1998)によれば極めて多くの部位が単語の読みに関与している。図１にその概要を示した。

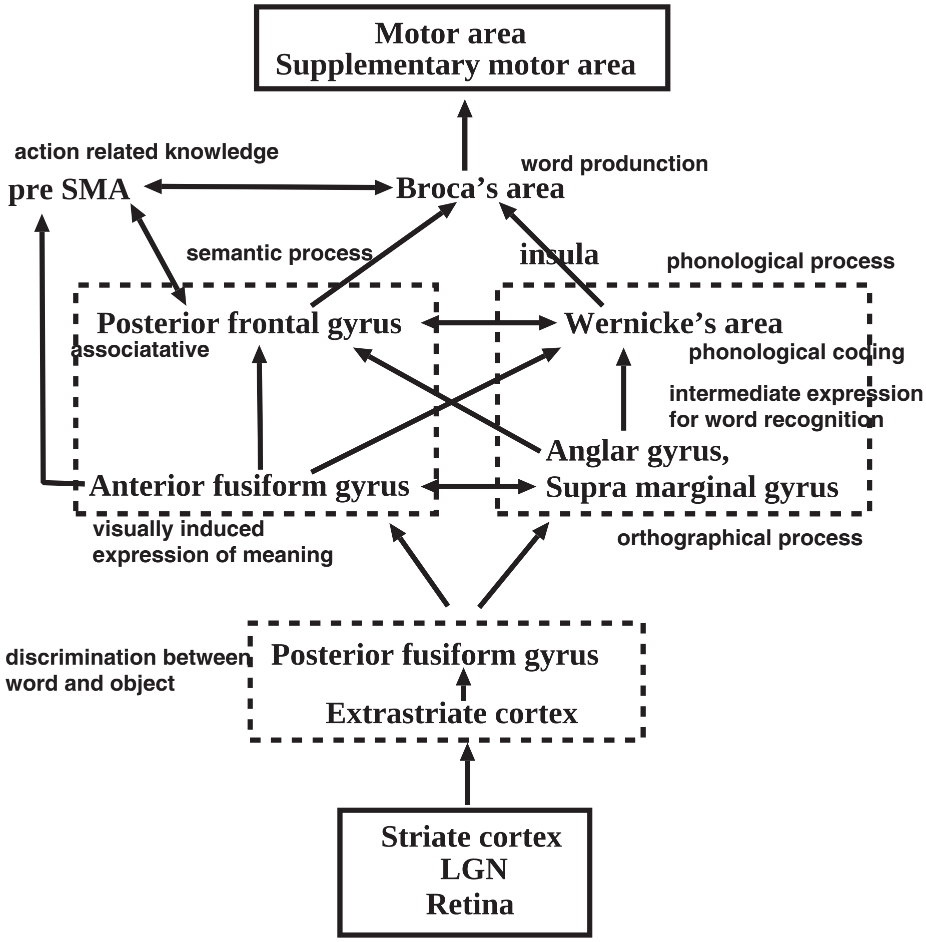

図１：単語の読みに関与する脳内部位の概念図

図では、左側に意味系の関与する部位を、左側に音韻系の関与する部位を示してある。

二重経路モデルとトライアングルモデル

ここで、規則語(regular words)と例外語(exception words)との区別、一貫語(consistent words)と非一貫語(inconsistent words)との区別をしておく。規則語とは書記素音韻対応規則に合致する単語である。書記素音韻対応規則に合致しない単語を例外語(たとえば yacht)と呼ぶ。hint, mint, saint, feint, lint などは発音記号/int/を共有しているので一貫語と呼び、一方 pint(発音記号は/páint/)は非一貫語と呼ばれる。例外語とは、その定義からして非一貫語でもある(Glushko, 1979, p.676)。すなわち、規則—不規則の概念と、一貫—非一貫の概念は独立ではない。

二重経路モデルは失読症(alexia)、あるいは難読症(dyslexia)と呼ばれる脳に障害を持つ患者のデータから考案された神経心理学的モデルである。表層失読症(surface dyslexia)患者は、規則語を読むことができるし、非単語も読むことができるが、例外語(特に低頻度の)を読むことが困難である。加えて、音韻失読症(phonological dyslexia)という症例では、実在する単語の読みは正常に保たれているにもかかわらず、非単語が読むことができない。この二重に乖離した現象を説明するために二重経路モデルが提案された。二重経路モデルは上記の神経心理学的データを素朴に表現した説明である。すなわち語彙経路に障害が発生すれば表層失読症が生じ、反対に書記素音韻対応規則経路の損傷は音韻失読症が表出する。

一方、トライアングルモデルでは以下のように失読症を説明する。表層失読は単一経路での障害のされ方によるとされる(Plaut, 1996, シミュレーション4)。音韻失読については、トライアングルモデルにおいて、一旦書記素へ入力された文字は、音韻に直接出力を送る直接経路と、意味を介する間接経路との両方の影響を受ける。ある単語がどちらの経路をたどって読まれるかは、単語ごとに異なるとされる。直接経路では規則語と高頻度例外語とが学習され、低頻度例外語は間接経路である意味層のサポートを必要とする。意味系からの関与の程度を労働の分割(division of labor)という。ただし、トライアングルモデルでは労働の分割問題を実装しているわけではない。トライアングルモデルでは書記素層から音韻層への直接経路だけがシミュレートされただけであり、モデルで説明できない部分を労働の分割と読んでいるだけである。

単語の集まりを語彙(lexicon)と呼ぶが、多くの理論では脳内にこの語彙辞書が存在すると仮定している。語彙辞書の分散表現を仮定するのがニューラルネットワークの発想である(Seidenberg & McClelland, 1989)。ニューラルネットワークを用いる言語研究における基本的なアイデアは単語は3つの要素(あるいは脳内の部位と言っても良い)に分散表現されていると仮定することである。3つの要素は書記素、音韻、意味と呼ばれる。書記素とは視覚提示された単語の脳内表現を指し、音韻とは音声単語の脳内表現、意味とは視覚

書記素音韻変換過程に関与する意味系の貢献を評価するためのモデル

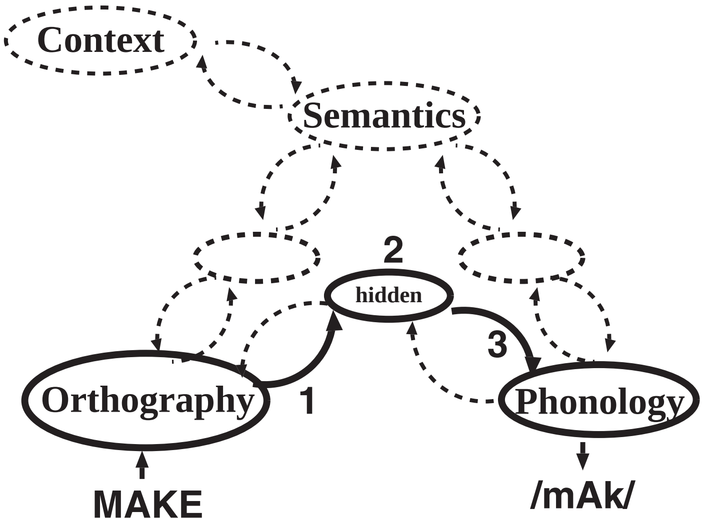

図２：トライアングルモデル。書記素から音韻への直接経路は３層の、
意味経由の間接経路は５層のネットワークと見なせる

や聴覚などの感覚入力に依存しない抽象的な表現である。

英語圏の人たちの間では、先天的に文字を読むことが困難な先天性（発達性）難読症（developmental dyslexia）と呼ばれる症状を示す人が人口の 20% にまで達すると言われている（Shaywitz, 1996）。Paulesu et al. (2001）によれば、印刷された文字（書記素）とその読み（音韻）との間にほぼ完全な一対一関係が存在するイタリアにおける難読症の発生率は英語圏の国の発症率の半分以下であるという。英語やフランス語は不規則語が存在するので準規則的（quasi-regular）な書記素系を持つ言語、あるいは深層書記素を持つ言語であると言われる。一方 Paulesu et al. (2001）によれば、例外語がほとんど存在しないイタリア語は浅層書記素言語と呼ばれる。

Paulesu, Frith, & Frackowiak (1993）は PET や fMRI を用いた実験で、難読症患者の賦活部位と統制群（健常者）の賦活部位に差が見られなかったと報告している。彼らの結果によれば、難読症患者と健常者との差位は、それらの部位の活性化の程度である。健常者の方が一貫して活性値が高かった。英語やフランス語のような、書記素と音韻との間に一定の規則性とその規則に従わない特殊な読みが存在する深層書記素（deep orthography）を持つ言語と、書記素と音韻の組合せに明確な規則が存在し、例外がほとんどない表層書記素（transparent or shallow orthography）を持つ言語との差位が、難読症の発生率と関係し、上記３つの言語間で、書記素から音韻への変換処理過程に関与する脳内部位は同一であることから、書記素を音韻へと変換する処理過程には、特定の言語によらない普遍的な処理

機構が存在し、その処理機構の活性化の低下が難読症を引き起こしている可能性があると考えられる。また、難読症はいわゆる知的な能力と無関係であると言われている。たとえば、トーマス・エジソン、アルバート・アインシュタイン、ウィンストン・チャーチル、アガサ・クリスティ、ジョージ・ワシントンなどが難読症であったという(Shaywitz, 1996)。

ここで、英語に代表される欧米の深層書記素を持つ表記体系の言語と、日本語の場合との比較を考えてみる。英語ではおなじ a でも make と cat では発音が異なる。a がどのように発音されるかはその単語の前後の文字によって定まる。その点、日本語の場合の平仮名やカタカナは表音文字であるので英語のような問題が無い。日本語の仮名はモーラを単位とする表記法であり、文字と音との対応関係に、少数の例外を除いて 1 対 1 対応が存在し、前後の文脈によって影響されることがほとんどない。一方、漢字については先の英語例と同じように単一文字だけでは読みが決まらない場合が多く、 1 つの文字が前後の文脈によりいく通りもの発音が存在する。日本語については、さらに、意味系の関与と視覚入力の関与に関しての研究として Iwata (1986) などがある。Iwata (1986) は、音韻的特徴と意味的特徴とを統制した刺激語を用いた。従って、被験者の反応は刺激語の形態的類似性によって生じた誤りだと考えられる(‘鍵’を‘銀’、‘耳’を月と読み間違えるなど)。対照的に、これらの患者は視覚照合課題の成績は正常の範囲だった。このことから、これらの患者は、視覚系入力を音韻情報へ変換することだけができなかったと考えられる。意味系の関与と視覚系との関与を、このような神経心理学的実証データと突き合わせることによって、より実り多い研究が期待できよう。

英単語は、仮名と漢字の中間に位置すると考えられる。すなわち文字と音との間の対応関係が規則的である規則語と規則に従わない不規則語、非一貫語とが存在するからである。

トライアングルモデルは直接経路と意味層を介した間接経路という 2 つのネットワークを持つニューラルネットワークモデルとみなすことができる。そして、トライアングルモデルにおける労働の分割問題は、書記素と意味とによる競合の結果とみなすことができる。この競合の過程を直接モデル化して数値として表現できれば、意味系の役割、各単語毎の意味系からの貢献の度合いを評価することができる。先述のとおり Plaut et al. (1996) では、意味経路は実装されなかった。論文中では推定上の意味経路(putative semantic pathway)を用いたとしか書かれていない。すなわち図 2 中の実線部分だけがシミュレートされただけである。浅川(2005)によれば、書記素と意味との両者の違いとは、競合する 2 つのネットワークの空間の領域分割の大きさという量的な問題に帰結される。ここで、あらためてニューラルネットワークモデルに単語の読みを学習させることの意味を

考えてみる。ニューラルネットワークに単語の読みを学習させるということは、書記素から音韻への変換規則を学習させることである。低頻度例外語にエラーが大きいのは、他の大部分の単語に共通する書記素音韻対応規則を学習し、その結果を適用しているからであって、ニューラルネットワークの見地からすれば正しい一般化と解釈することができる。すなわち、未学習のデータに対して、学習によって獲得した書記素音韻対応規則を適用しているという意味である。むしろ、高頻度例外語は学習のしすぎ、すなわち過学習なのである。例外語と規則語との頻度効果の交互作用は、このような説明が可能である。そして、直接経路と間接経路との労働の分割問題は２つのネットワーク間の競合あるいは協調作用とみなすことができるだろう。もし、書記素音韻対応規則を学習し、例外語と規則語を自動的に分類して学習できるアルゴリズムが存在すれば、表層失読と音韻失読という神経心理学的症状を説明するための労働の分割問題を解決できるモデルとなる。

LDM モデル

問題点は先に指摘したとおり、二重経路モデルにおける２つの経路を融合するパラメータを探索的に手で調整していること、トライアングルモデルにおいては意味系の関与の程度を測り、かつ実装するモデルがないことである。英単語の読みを扱ったモデルが説明しなければならないことは、二重経路モデルにおいては語彙経路と書記素音韻対応規則経路との融合をどのように実装するか、あるいはその言い替えであるが、トライアングルモデルでは労働の分割問題をどのようにして実現するのか、という問題を解決することである。本稿では、書記素音韻対応規則を学習し、同時に規則語と例外語とを自動的に分類して学習できるモデルを提案する。このモデルとその計算アルゴリズムは、混合エキスパートネットワークモデル（Mixtures of experts model; Jacobs, Jordan, Nowlan, & Hinton,1991; Jordan & Jacobs, 1994）に基づくものである（図３）。各エキスパートは単純なフィードフォワード型のネットワークであり、すべてのエキスパートは同じ入力を受け取り同じ数の出力ユニットを持っている。ゲーティングネットワークもフィードフォワード型のネットワークであり、エキスパートネットワークと同じ入力を受け取る。図中の g はゲーティングネットワークの出力（確率）を表し、すべてのゲーティングネットワークの和は１となる。u はエキスパートの出力である。エキスパートネットワークの出力はゲーティングネットワークによって重み付けられた合成変量となる。以降、本稿で提案するモデルを LDM（Labor Division Model）と表記する。

定式化

観測された入力データを $x = x_1$ とし出力データを y とする。LDM は、図４に示すよう

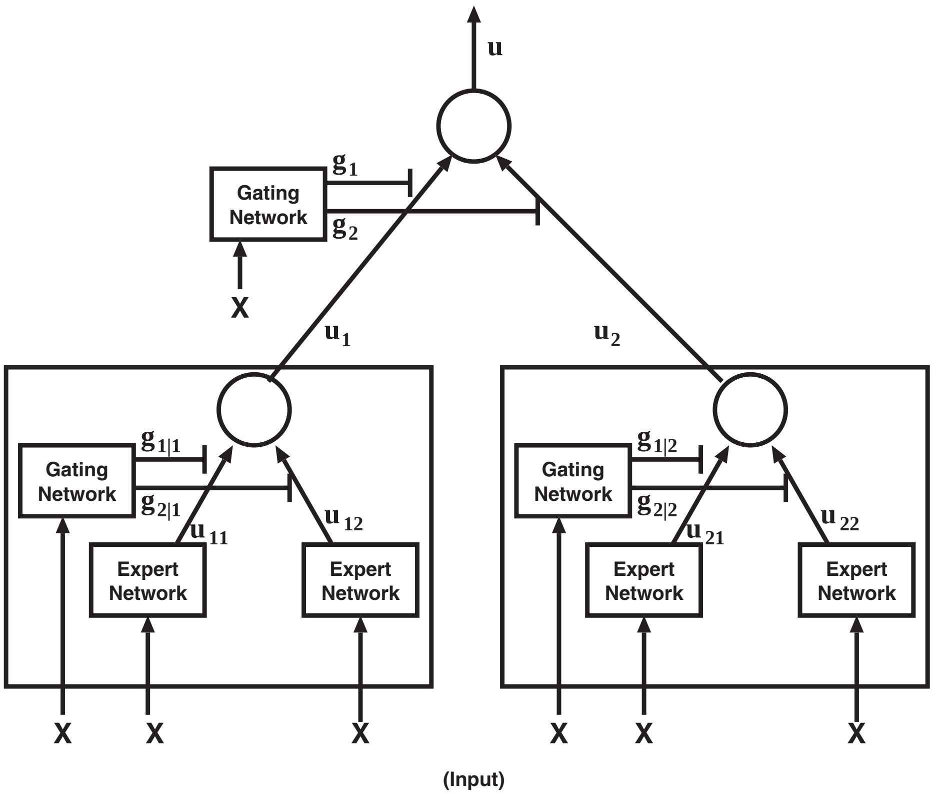

図３：２段階の混合エキスパートネットワークモデルの概念図

に２つの下位ネットワークを持つ。それぞれ、書記素と意味の情報を処理するネットワークである。書記素ネットワークは入力ベクトル $x=x_1$ に対して出力ベクトル y_1 を出力する。一方、意味ネットワークは入力ベクトル $x_2=m(x)$ に対して出力ベクトル y_2 を出力する。ここで、$m(x)$は、書記素入力を意味への入力ベクトルへ変換する任意の関数とする。この $m(x)$ は、図１では後紡錘状回(posterior fusiform gyrus)から前紡錘状回(anterior fusiform gyrus)へ至る経路における変換に相当する。また、図２においては、書記素から意味へ至る変換過程に相当する。各々の出力ベクトル y_i は、ゲーティングネットワーク g_i によって混合される。

下位ネットワークとは、入力ベクトル x_i を出力ベクトル y_i へと変換するためのリンク関数である。リンク関数とは、入力ベクトル x_i に対して、入力ベクトル x_i を出力ベクトル y_i へ結びつける関数という程度の意味である。リンク関数は問題によってどのような関数を用いてもよい。２層のパーセプトロンすなわち、単純な線形変換でもよいし、各要素に対してシグモイド関数を用いるような非線形変換でも、あるいはガウシアン関数でも良い。本稿では、リンク関数として最も単純な線形変換、すなわち入力層と出力層を持つ

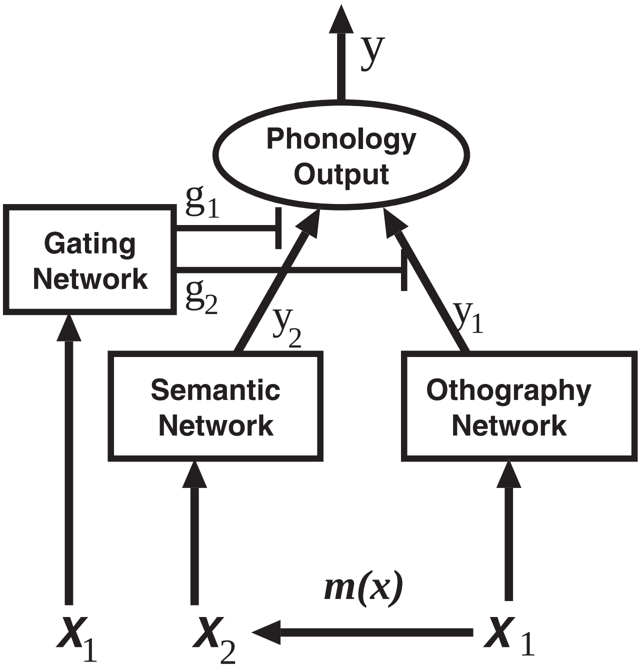

図４：LDM の模式図

単純な２層のパーセプトロンを用いた。

ゲーティングネットワークは１つの数値(確率とみなすことができる)を出力する。ゲーティングネットワークも線形関数であり x を変数とする以下の式のように表される。

$$\xi_i = v_i^T x \quad (1)$$

ここで ν_i は結合係数ベクトルであり、右肩の添字 T はベクトルの転置を表している。ゲーティングネットワークの出力は、ゲーティングネットワーク関数の出力 ξ_i を用いた次のソフトマックス関数として定義される。

$$g_i = \frac{\exp\{\xi_i\}}{\sum_k \exp\{\xi_k\}} \quad (2)$$

g_i は正であり、すべての g_i について合計すると必ず１になる。すなわちソフトマックス関数は入力空間をその値によってどのように分割するかを示す指標である。式(2)において $k = 2$ の場合を考えれば、簡単な変形によって g_i はシグモイド関数となることがわかる。

$$g_1 = \frac{\exp\{\xi_1\}}{\exp\{\xi_1\} + \exp\{\xi_2\}} \quad (3)$$

$$= \frac{1}{1+\exp\{-(v_1-v_2)^T x\}} \tag{4}$$

であり$(\nu_1-\nu_2)^T x$を引数とするシグモイド関数となる。$g_1+g_2=1$となることは自明であり、２番目のゲーティングネットワークの出力g_2は$1-g_1$に等しい。この関数はベクトル$\nu_1-\nu_2$と入力データベクトルxとの内積によって定義されるシグモイド関数であり、$\nu_1-\nu_2$は入力空間をどのように分割するかを示す。

$$g_1=g_2=\frac{1}{2} = \frac{1}{1+\exp\{-(v_1-v_2)^T x\}} \tag{5}$$

$$-(v_1-v_2)^T x = 0 \tag{6}$$

という計算により$\nu_1-\nu_2$に直交するベクトル上の点の集合では、２つの下位ネットワークの影響は等しくなる。このベクトル付近の空間の分割の滑らかさの度合は$\nu_1-\nu_2$というベクトルの長さの絶対値によって定まる。このベクトルと入力データベクトルxとの内積の値によって、一方の下位ネットワークの出力が他方の出力よりも重視されることになる。式(4)中の$\nu_1-\nu_2$の絶対値が小さければ$\nu_1-\nu_2$のまわりのベクトルで定義されるシグモイド関数はなだらかになり、ソフトな空間の分割、すなわち２つの下位ネットワークの出力はある割合で混合されて、最終的な解を出力することになる。この$\nu_1-\nu_2$の絶対値が０に近づくとき、したがって$g_1=g_2=1/2$に近づくとき、いかなる入力ベクトルxに対しても、両ネットワークは等しく貢献することになる。反対に$\nu_1-\nu_2$の絶対値が大きければ空間を分割するシグモイド関数はステップ関数に近づき、空間を明確に分割する役割を果たす。

まとめると、あるx_1とx_2とが与えられると全体の出力yは$g_1y_1+g_2y_2$となる。すなわち下位ネットワークの重みづけ平均であり、空間の分割は、$\nu_1-\nu_2$というベクトルの向きによって定まる。

LDMの確率的解釈

LDMを確率的に解釈することができる。上述のとおり、データxをyへ写像する関数と、その関数の出力を調節、あるいは混合するためのゲーティングネットワークがある。x_1すなわち書記素の入力について、g_1の値を、意思決定に関わる確率と解釈する。同様にして、x_2を意味の入力、g_2をその意思決定に関わる確率と解釈する。下位ネットワークの出力y_iを、入力データx_i、とデータの密度関数を表すパラメータθ_iとを用いて条件付確率として定式化することができる。

書記素音韻変換過程に関与する意味系の貢献を評価するためのモデル

$$P(y_i|x,\theta) = \sum_{i=1}^{n} g_i(\{v_i\},x_i)\, p_i(y_i|x_i,\theta_i) \tag{7}$$

ここで θ_i は分布の密度関数を定めるパラメータベクトルである。

$$\theta_i = \begin{bmatrix} \mu_i \\ \Sigma_i^2 \end{bmatrix} \tag{8}$$

のような多次元正規分布を考えれば、平均が μ かつ、共分散行列 Σ が $\sigma^2 I$、(I は単位行列)で与えられるとすると、

$$p_i(y_i|x_i,\theta_i) = \frac{1}{\left(2\pi\sigma_i^2\right)^{n/2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_i-\mu_i)^T(y_i-\mu_i)\right\} \tag{9}$$

と表すことができる。最終的な出力 y の確率は式(7)と式(9)を用いて、

$$P(y|x,\theta) = \sum_i g_i \frac{1}{\left(2\pi\sigma_i^2\right)^{n/2}} \exp\left\{-\frac{1}{2\sigma_i^2}(y_i-\mu_i)^T(y_i-\mu_i)\right\}. \tag{10}$$

この値は、g を既知とした、ガウシアン密度関数によって生成されるモデルである。ここで分散パラメータ σ_i^2 は μ_i を中心とする超円の大きさ(半径)を定めるものと解釈することができる。$\sigma^2 \rightarrow 0$ の極限ではディラック(Dirac)のデルタ関数 δ となる。ディラックのデルタ関数は、

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1, \tag{11}$$

であり、かつ、

$$\begin{cases} \delta(x) = \infty, & x = 0 \text{ のとき} \\ \delta(x) = 0, & x \neq 0 \text{ のとき} \end{cases} \tag{12}$$

を満たす関数であるから、$\delta(x)$ は 1 点 $x = 0$ を除いていたるところ 0 である。$x = 0$ では無限大の値をとり $x = 0$ を含む区間上の積分が 1 となる。ディラックのデルタ関数の定義は複数存在するが、その中の一つとして正規分布の分散 σ^2 を 0 に近づけたときの極限 $\lim_{\sigma^2 \rightarrow 0} N(0,\sigma^2)$ としても定義できる並木(1982)。ここで N は正規分布の確率密度関数である。

学習アルゴリズム

誤解のない限り、$p_i(y_i|x_i,\theta_i)$ を p_i と表記する。ベイズの定理を使って事後確率を定義すると

$$h_i = \frac{g_i p_i}{\sum_k g_k p_k} \tag{13}$$

データ集合 $\left\{(x^{(t)}, y^{(t)}\right\}_{t=1}^{N}$ について対数尤度 l を次のように定義する:

$$l = \mathcal{L}\left(\theta_i; x^{(t)}, y^{(t)}\right) \;=\; \ln \prod_{t=1}^{N} \prod_{i=1}^{2} g_i^{(t)} p_i^{(t)} \tag{14}$$

$$= \sum_{t=1}^{N} \sum_{i=1}^{2} \ln\, g_i^{(t)} p_i^{(t)} \tag{15}$$

従って、

$$l = \sum_{t=1}^{N} \sum_{i=1}^{2} \left[\ln\, g_i^{(t)} \frac{1}{\left(2\pi\sigma_i^2\right)^{n/2}} \exp\left\{ -\frac{1}{2\sigma_i^2} \left(y_i^{(t)} - \mu_i\right)^T \left(y_i^{(t)} - \mu_i\right) \right\} \right] \tag{16}$$

この式を μ_i について微分することで、

$$\frac{\partial l}{\partial \mu_i} \;=\; \sum_{t=1}^{N} \sum_{i}^{2} g_i \left[\frac{p_i\, \exp\left\{ -\frac{1}{2\sigma^2} \left(y_i^{(t)} - \mu_i\right)^2 \right\}}{\sum_j p_j\, \exp\left\{ -\frac{1}{2\sigma^2} \left(y_j^{(t)} - \mu_j\right)^2 \right\}} \left(\frac{y_i^{(t)} - \mu_i}{\sigma_i^2} \right) \right] \tag{17}$$

$$= \sum_{t=1}^{N} \sum_{i}^{2} g_i \left[h_i \left(\frac{y_i^{(t)} - \mu_i}{\sigma_i^2} x \right) f'(x) \right] \tag{18}$$

を用いて勾配降下法を使って学習を行なうことができる。すなわち、各ネットワークとして単純な 2 層のパーセプトロンを用いれば、複雑な学習則を用いなくても近似解を求めることができると予想される。またプログラムを書くのも容易となる。μ_i の更新式は、

$$\Delta\mu_i = \rho \sum_{t=1}^{N} h_i \frac{\left(y_i^{(t)} - \mu_i\right)}{\sigma_i^2} x_i^{(t)} \tag{19}$$

となる。ここで ρ は学習係数である。σ の学習も同様にして、l を σ で微分した値を用いて学習を行うことができる。p_i を σ で微分すると、

$$\begin{aligned} \frac{\partial p_i}{\partial \sigma_i} &= \frac{1}{(2\pi)^{n/2}\sigma_i^n} \exp\left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right\} \\ &= -n\, \frac{1}{(2\pi)^{d/2}\sigma^{n+1}} \exp\left\{ -\frac{1}{2} \frac{(y_i - \mu_i)^2}{\sigma_i^2} \right\} \end{aligned} \tag{20}$$

$$+\frac{1}{(2\pi)^{(d/2)}\sigma_i^n}\frac{|y_i-\mu_i|^2}{\sigma_i^3}\exp\left\{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma_i^2}\right\} \tag{21}$$

$$= -\frac{d}{\sigma_i}\frac{1}{(2\pi)^{d/2}\sigma^n}\exp\left\{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma_i^2}\right\}$$

$$+\frac{1}{(2\pi)^{(d/2)}\sigma_i^n}\frac{|y_i-\mu_i|^2}{\sigma_i^3}\exp\left\{-\frac{1}{2}\frac{(y_i-\mu_i)^2}{\sigma_i^2}\right\} \tag{22}$$

$$= \frac{n}{\sigma_i}p_i+\frac{|y_i-\mu_i|^2}{\sigma_i^3}p_i \tag{23}$$

$$= \left\{\frac{|y_i-\mu_i|^2}{\sigma_i^3}-\frac{n}{\sigma_i}\right\}p_i \tag{24}$$

それゆえ $\partial l/\partial\sigma_i$ は

$$\frac{\partial l}{\partial\sigma_i} = \sum_{t=1}^{N}\frac{1}{\sum_i g_iP_i}\frac{\partial}{\partial\sigma_i}\sum g_iP_i \tag{25}$$

$$= \sum_{t=1}^{N}\frac{g_iP_i}{\sum_k g_kP_k}\left\{\frac{d}{\sigma_i}+\frac{\left|y^{(t)}-\mu_i\right|^2}{\sigma_i^3}\right\} \tag{26}$$

$$= \sum_{t=1}^{N}h_i\left\{\frac{\left|y^{(t)}-\mu_i\right|^2}{\sigma_i^3}-\frac{d}{\sigma_i}\right\} \tag{27}$$

となるので、σ_i の更新式は

$$\Delta\sigma_i=\rho\sum_{t=1}^{N}h_i\left\{\frac{\left|y^{(t)}-\mu_i\right|^2}{\sigma_i^3}-\frac{d}{\sigma_i}\right\} \tag{28}$$

となる。

次に、ゲーティング関数の引数である ν_i の更新式は、次のようになる。g_i を ν_i で微分すると、

$$\frac{\partial g_i}{\partial v_i} = \frac{\partial}{\partial v_i}\frac{\exp\{v_ix\}}{\sum_k\exp\{v_kx\}} \tag{29}$$

$$= \frac{x\exp\{v_ix\}\sum_k\exp\{v_kx\}-\exp\{v_ix\}x\exp\{v_ix\}}{\{\sum_k\exp\{v_ix\}\}^2} \tag{30}$$

$$= \frac{\exp\{v_ix\}\sum_k\exp\{v_kx\}-\{\exp\{v_ix\}\}^2}{\{\sum_k\exp\{v_ix\}\}^2}x \tag{31}$$

$$= \left[\frac{\exp\{v_i x\}\sum_k \exp\{v_k x\}}{\{\sum_k \exp\{v_i x\}\}^2} - \frac{\{\exp\{v_i x\}\}^2}{\{\sum_k \exp\{v_i x\}\}^2}\right] x \tag{32}$$

$$= \left(g_i - g_i^2\right) x \tag{33}$$

をえる。したがって、$\partial l / \partial \nu_i$ は

$$\frac{\partial l}{\partial v_i} = \sum_{t=1}^{N} \frac{1}{\sum_i g_i P_i} \frac{\partial g_i}{\partial v_i} \sum g_i P_i \tag{34}$$

$$= \sum_{t=1}^{N} \frac{1}{\sum_i g_i P_{ij}} g_i (1 - g_i) x^{(t)} P_i \tag{35}$$

$$= \sum_{t=1}^{N} \left[\frac{g_i P_i}{\sum_i g_i P_i} - g_i \frac{g_i P_i}{\sum_i g_i P_i}\right] x^{(t)} \tag{36}$$

$$= \sum_{t=1}^{N} (h_i - h_i g_i) x^{(t)}, \tag{37}$$

となる。ν_i の更新式は

$$\Delta v_i = \rho \sum_{t=1}^{N} (h_i - h_i g_i) x^{(t)} \tag{38}$$

となる。

LDM による英単語の読みの説明

トライアングルモデルは直接経路と間接経路という 2 つの下位ネットワークを持つ LDM とみなすことができる。LDM において出力が調節されることと労働の分割問題は直接対応すると考えられる。すなわち、トライアングルモデルにおける労働の分割問題は LDM におけるゲーティングネットワークによる領域の分割である。

$\sigma^2 \to 0$ の極限ではディラックのデルタ関数となり、特定の単語にしか応答しないニューラルネットワークを作ることができる。逆に σ^2 を大きくすれば多くの英単語に応答するニューラルネットワークモデルになる。前者を低頻度非一貫語、後者を一貫語、高頻度非一貫語と見なしうると考える。すなわち、σ^2 が小さければ、非一貫語(たとえば pint)だけに選択的に応答し、その近隣に付置されていると考えられる一貫語(たとえば hint,mint, print, lint)には、まったく応答しない英単語の読みのモデルを実現できる。英単語書記素の羅列からなる多次元空間を想定し、その空間内のただ一点のみを指定する関数であるディラックのデルタ関数をもちいて、二重経路モデルにおけるルックアップテーブルを実装したと捉えることも可能である。一方、分散パラメータ σ^2 が大きければ、複数の単語に

わたって一貫した書記素音韻対応規則が実現されたものとみなすことができる。

LDMによって、ゲーティングネットワークの存在とエキスパートネットワークの存在から、多くの単語と読みを共有する一貫語が特定の下位ネットワークによって集中的に処理され、頻度の少ない非一貫語が意味の経路の貢献を必要とするという説明が可能になると思われる。さらに、頻度効果についてもある程度説明可能ではないかと考える。

考　察

本稿では、文字を読む際の脳内過程の計算論的モデルとして、LDMを提案した。欧米諸語の表記体系と漢字と仮名との2種類の表記体系とを持つ日本語の表記体系とでは、同列に語ることは困難であると考える。書記素と音韻との間に1対1対応が存在する仮名では、表層書記素体系であるイタリア語などに類似したデータが得られると予想され、一方、書記素と音韻との間に複数の対応関係が存在する漢字の読みに関しては、深層書記素体系である英語に近い反応が予想できる。このような予想は目新しいものではない(たとえばIwata, 1986)。そして、単語の読みに関して意味系の関与が必要となる深層書記素系の言語、あるいは、漢字の読みに関しては、そのモデルが提案されていた(Coltheart et al., 2001; Plaut et al., 1996)ものの、ある特定の単語がどの程度の意味系の関与を受け、また別の単語がどの程度であるのかを定量的に評価するモデルが存在しないというのが現実であった。そこで本稿では、この問題を解決するためのモデルを考察した。モデルの応用可能性は、英語、フランス語、及び、日本語の漢字などの深層書記素系、そして、イタリア語、ハングル、日本語の仮名などの表層書記素系、いずれにも適用可能である。この意味において、各言語への比較文化的視点を提供できる可能性のあるモデルであると考える。

さらに、バイリンガル話者、マルチリンガル話者の発話モデルとしての応用可能性が指摘できるかも知れない。本稿では意味系と書記素系という2つの処理系のみを考えてきた。数式上の添字は一貫して$i=\{1,2\}$とした。特に⒁式以降、iが2までしか無いことを明記した。しかし、当然ながらLDMは$i>2$に拡張できる。この拡張によって複数の文化的背景を持つ多様な社会科学的データを分析するために役立つかもしれない。

二重経路モデルが二つの経路の融合パラメータを手で探索的に定めなければならないこと、トライアングルモデルが労働の分割問題の実装を示していないことから、二重経路モデルやトライアングルモデルがどれだけ人間の成績に近づこうとも、本質的な解決にはならないと考える。モデルの優劣をきめるのは、課題成績だけではなく、その背後にあるモデルの持つ本質、換言すれば心理学モデルとしての実現可能性を議論しなければならない。

ある程度万能なアルゴリズムを用いれば任意の精度で望む関数が実現できることはノーフリーランチ定理が主張しているところでもある Duda, Hart, Peter, & Stork (2001)。この“ただの昼飯はない定理”とは、どんなにうまく学習アルゴリズムを選んでも、もし、すべての目標関数の存在可能性が等しければ、良いアルゴリズムが悪いアルゴリズムに勝るとは言えないことを主張する定理である。さらに、どんなアルゴリズムを採用しようとも、ランダム推測の方が優れている目標関数が少なくとも一つ存在する。与えられた課題に独立などんな学習アルゴリズムの優劣判定方法はない。どのようなモデルが最も優れているのかは、問題の種類や、付随する情報によって決まる。この意味において、LDM は労働の分割問題を自然に取り入れた実装になっており、しかも二重経路モデルにおけるルックアップテーブルの存在も説明可能なことから、同じ成績を示すのであれば最も自然なモデルだと言うことができると考えられる。

トライアングルモデル Plaut et al. (1996) は書記素に入力された情報を中間層を経て音韻に変換するためのニューラルネットワークであり、この意味で 3 層パーセプトロンである。計算論的には 3 層パーセプトロンの能力は、般化能力によって測定されるべきである。般化能力の観点から見れば、高頻度不規則語の成績が優れているのは過学習 (over-learning) であるとみなすことができる。一方、未知の単語の規則語を正しく読めるという事実はトライアングルモデルが一般化能力を有していると見なすことができる。このように高頻度規則語については過学習であり、規則語に対しては一般化可能な程度の学習をさせたことに相当し、どこで学習を打ち切るかに依存して、トライアングルモデルは如何様にも成績が変化する可能性を持っていると考えられる。Plaut et al. (1996) はこのような学習打ち切り基準について何も述べていない。一方、LDM はその学習アルゴリズムから見て、過学習が起こりにくいと思われる。例外的な事例は、別のゲーティングネットワークによって処理されるように訓練され、多くの英単語によって定義される書記素音韻対応規則とは別のエキスパートネットワークとして表現されるようになると考えられるからである。

語彙経路、あるいは間接経路を介したネットワークでは、例外語に特化し、局在化した領域分割がなされていると考えれば、語彙経路、労働の分割問題が説明できる。分散パラメータを $\sigma^2 \rightarrow 0$ とした極限を考えれば、ある単語の占める領域はディラックのデルタ関数に等しくなり、この各々の例外語ごと、その単語のみに応答するルックアップテーブルと同一視できる。すなわち LDM の観点から二重経路モデルとトライアングルモデルを解釈し直すとすれば、なぜ例外語は語彙経路、あるいは間接経路を通るのかを見通しよく説明することができる。この意味において、二重経路モデルとトライアングルモデルの間に本質的な違いはない。両者の間に存在する違いとは、分散パラメータによる入力空間の領

域分割の大きさという量的な相違に換言される。二重経路モデルとトライアングルモデルとは、LDMという、より一般的で、より包括的なニューラルネットワークモデルの特殊な場合として、本質的には同じ範疇のモデルに属すると言えよう。

Plaut et al.(1996)の論文もColtheart et al.(2001)の論文も引用数の多さから見れば、非常に有力なモデルであることは確実である。Plaut et al.(1996)の論文は、命名課題における頻度効果、規則性効果(および両者の交互作用)などを反応時間の側面から検討し、さらに失読症の事例の詳細なシミュレーションが示されている。一方、Coltheart et al.(2001)のそれは、Plaut et al.(1996)のモデルや他のモデルと比較しながら、命名課題における18種類の効果、語彙決定課題における5種類の効果、失読症などに関する詳細はシミュレーションが示されている。従って、本稿で提案したLDMが二重経路モデルに比べて優れていると言うには無理がある。むしろ、これから解決すべき点が多い未完成のモデルである。しかし、LDMを用いることで二重経路モデルとトライアングルモデルとでは記述が困難であったり、実現方法が不明瞭であった問題点を、明確に定式化し、実装が可能であることは示すことができたと考える。

最後に、本論文のはじめに触れた頻度効果の問題について考える。頻度効果を実装するためにPlaut et al.(1996)は、各々の単語の頻度の対数を学習時に結合強度の更新の際に掛け合わせるという形で実現した。しかし、誰でもが疑問に思うこととして、対数変換を用いることに心理学的意味を見出せない。LDMの枠組で頻度効果を説明するとすれば、以下のようになる。LDMにおいては、どちらの経路を重視するかが自動的に学習される。あるいは、LDMでは、どちらの経路をどの割合で混合するかが自動的に学習される。従って、一貫語、すなわち、他の単語と共通の読みを共有する単語は自動的に多数回訓練され、そのために結合強度はより一般的な書記素音韻対応規則を学習するようになる。この学習は、ゲーティングネットワークの出力選択によってより頻繁に起こることになるので、結果として一貫語を処理できるようになり、非一貫語は別の経路をたどることになるので、非一貫語の提示によって書記素音韻対応規則の学習に干渉することが少なくなると考えられる。

枚数の制約から具体的なシミュレーションや、詳細な分析、十分な考察ができなかったが、稿を改めて報告することとしたい。

要　約

本稿では英単語の読みのモデルである二重経路モデルとトライアングルモデルを統合し、二重経路モデルにおけるルックアップテーブルとトライアングルモデルにおける労働の分割問題を解決するためのモデルLDMを提案した。ガウシアン関数に基づくLDMの

確率的解釈を導入することで、ゲーティングネットワークによって、その特殊な場合として、ガウシアン関数の分散パラメータ σ^2 を 0 に近付けて、デルタ関数のようにして、二重経路モデルの仮定している例外語のルックアップテーブルを説明可能である。そして、ゲーティングネットワークによる空間分割によって、特定の単語だけが別の経路をたどり、他の単語は書記素音韻対応規則にしたがって処理されるかを LDM によって表現可能である。このことは労働の分割問題に対する一つの解を与えているように思われる。LDM による説明を用いれば、二重経路モデルとトライアングルモデルの間に本質的な違いはなく、両者の間に存在する違いは、分散パラメータによる入力空間の領域分割の大きさという量的な問題に帰結されることになる。すなわち、LDM を用いることで二重経路モデルとトライアングルモデルとでは記述が困難であったり、実現方法が不明瞭であった問題点を、明確に定式化し、実装可能であり、両モデルをその特別な場合として含む、より一般的で、より包括的なモデルの一部として、統合して説明できる可能性があると思われる。

謝　辞

本稿をまとめるにあたり、助力を頂いた岩船幸代氏に感謝する。彼女の助力がなければ本稿は完成しなかった。

引用文献

浅川伸一.（2005）. 英単語の読みにおける二重回路モデルとトライアングルモデルの統合の試み.『心理学研究』, *75*, 523-529.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distriputed-processing approaches. *Psychological Review, 100*, 589-608.

Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1197-1211.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). Drc: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Duda, O., Hart, I., Peter, E., & Stork, D. G. (2001). *Pattern classification, 2nd edition*. New York: John Wiley and sons.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psyhology: Human Perception and Performance, 5*, 674-691.

Iwata, M. (1986). Neural mechanism of reading and writing in the japanese language. *Functional Neurology, 1*, 43-52.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*, 79-87.

Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixtures of experts and the em algorithm. *Neural Computation, 6*, 181-214.

McClelland, J. L., & Rumelhart, D. E.(1981). An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review, 88*, 375-407.

Morton, J. (1980). The logogen model and orthographic structure. In U. Firth (Ed.), *Cognitive processes in spelling*. Academic Press.

並木三喜雄. (1982). 『デルタ関数と微分方程式』. 東京：岩波書店.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding in mind by simulating the brain*. MA: MIT Press.

Patterson, K., Plaut, D., McClelland, J. L., Seidenberg, M. S., Behrmann, M., & Hoges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In J. Reggia & E. Ruppin (Eds.), *Neural modeling of cognitive and brain disorders* (p.177-199). New York: World Scientific.

Paulesu, E., Démonet, J., Fazio, F., McCrory, E., Chanoine, V., Bruswick, N., Chappa, S., Cossu, G., Habib, M., Frith, C., & Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165-2167.

Paulesu, E., Frith, C., & Frackowiak, R. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342-345.

Plaut, D. (1996). Relearning after damage in connectionist networks: Toward a theory of rehabilitation. *Brain and Language, 52*, 25-82.

Plaut, D. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12*, 765-805.

Plaut, D., McClelland, J. L., & Seidenberg, M. S. (1995). Reading exception words and pseudowords: Are two routes really necessary? In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, & P. Cairns (Eds.), *Connectionist models of memory and language* (p.145-159). London: University College London Press.

Plaut, D., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. the contextual enhancement effect and some tests and extension of the model. *Psychological Review, 89*, 60-94.

櫻井芳雄. (1998). 『ニューロンから心をさぐる』. 東京：岩波書店.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Seidenberg, M. S., Plaut, D., Petersen, A. S., McClelland, J. L., & McRae, K. (1994). Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1177-1196.

Shaywitz, S. E. (1996). Dyslexia. *Scientific American, 275*, 98-104.

〔情報処理センター助手(計算論的神経科学・認知科学) 2010～12年度個人研究員〕