

異体字選好における単純接触効果と一般対応法則の関係

横山 詔一 (国立国語研究所)

ディスクリプタ：選好，単純接触効果，価値関数，
対応法則，フェヒナー法則，親近度

1. はじめに

言語生活はどの表現を選択するかという意思決定 (decision making) の連続である。方言と共通語の使い分けや、相手に対する呼称を「様—さん」のいずれにするかなど、ある表現が選択された背景には相手との親疎関係や上下関係などいろいろな要因が意識的・無意識的に影響している。それらのなかで重要な位置を占めるのは「選好 (preference)」であろう。人間は好きな表現を選んで使う傾向にあると考えられる。

では、日常生活における言語表現や言語行動の選択は、どのようなメカニズムによって決定されるのであろうか。日々の文字生活の中で、人間は意識的・無意識的にさまざまな文字刺激に接触している。その接触頻度の高低によって、その文字に対する記憶痕跡の強度が変化し、それが心的辞書 (mental lexicon) の形成や言語行動に影響を与える。図 1 は文字生活を捉えた俯瞰図である (横山, 2006a)。ここには示していないが、接触頻度の要因以外に、嘘字をきらったりする規範意識や字体差に注意を向ける傾向が生まれて、文字に対する「なじみ」や「好み」が左右される可能性もある。

今後、方言学や社会言語学においても、どの表現形式が選択されやすいのかを数量的に予測できる理論の構築と、検証データの積み上げが必要となるだろう。本研究は、社会心理学、精神物理学、動物心理学、経済学、医学統計学などの知見を参考にしながら、異体字の選択に関する新たなモデルを提唱した。

Mere Exposure Effect and Generalized Matching Law for Preference of Kanji Form.

YOKOYAMA Shoichi (National Institute for Japanese Language)

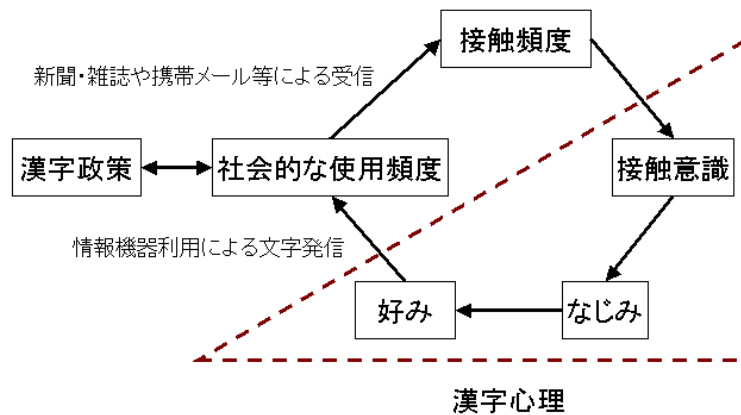


図1 文字生活の俯瞰図

選好に影響する要因の説明でよく知られているのが、社会心理学の分野で Zajonc (1968) が提唱してきた「単純接触効果 (mere exposure effect)」である。単純接触効果とは、なじみのない新奇な刺激に繰り返し接触しているだけで、その刺激に対する選好度が高まるという現象である。これは図1の「接触頻度 → 接触意識 → なじみ → 好み」の流れに相当する。Zajonc (1968) は単語の使用頻度と好意度 (favorability) の関係について調べる目的で、米国英語の各種辞書から「able – unable, better – worse, clean – dirty, good – bad, peace – war, life – death」などの対義語 154 ペアを抽出し、ペアのどちらの単語が好きかを2肢強制選択法 (2-Alternatives Forced Choice) で尋ねた。次に Thorndike & Lorge (1944) の語彙表を用いてペアのそれぞれに使用頻度のデータを付し、選好との関係を分析した。その結果、ペアのうち実験参加者に選好される単語は使用頻度の高い方であることが示唆された。例えば able と unable のペアでは実験参加者の 100% が able を選好した。使用頻度は able が 930 で unable は 239 であり、able の使用頻度の方が高かった。

Zajonc (1968) はさらに木 (pine など)、果物 (apple など)、野菜 (corn など)、花 (rose など) の単語を一つずつ実験参加者に呈示し、どの程度好きかを 0~6 の 7 段階評定尺度で尋ねた。その結果、単語の好意度と使用頻度の間には正の相関があることがうかがわれた。これらから、日常生活でよく使われる単語ほど好意度が高く、また好意度の高い単語ほど使用頻度が高くなるという解釈が導かれた。

さて、計量国語学や社会言語学の分野では、異体字選好の研究がいくつか行われている (横山, 2006a, 2006b ; 笹原・横山, 2000, 1998)。その手法は、異体字ペアを実験参加者

に呈示して「字体選好課題」を実施するというものである。字体選好課題とは「桧一檜」など 263 字種の新旧字体ペアを実験参加者に呈示し、ワープロやパソコンなど情報機器で字を書く場面をイメージしたときにより使いたいと感じる方の字体を 2 肢強制選択法で直観的に選ばせるという課題である。

さらに、「字体親近度比較課題」を用いた研究も存在する。字体親近度比較課題とは、新旧字体ペア 263 組のうち、字を見てより「なじみ」を感じる方の字体を 2 肢強制選択法で判断させるというものである。以下、なじみの主観的な感覚を「親近度 (familiarity)」という。字体の選好はおもに情報機器を使って書く行為に影響する要因で、親近度は読むことに関係する要因だと考えられる。

本研究は、新聞コーパス (corpus) を用いて計数した漢字頻度データから、字体選好課題の結果を予測するモデルを作成した。新聞における文字の出現頻度データは、それに人間が接触する確率を代表する一つの指標だと考えられる (横山・笹原, 2000)。予測モデルの導出は、精神物理学の分野で有名な「フェヒナー法則 (Fechner's law)」を出発点として、医学統計分野で発展が目覚ましい「ロジスティック回帰分析 (logistic regression analysis)」と動物行動学で Herrnstein (1961) や Baum (1974) が提唱した「一般対応法則 (generalized matching law)」を数理的に結びつける方向で進めた。これは横山 (2006b) の考えを発展させた理論であり、人間の意思決定のみならず、動物の選択行動や経済学での商品選択の解析にも応用可能な定式化を射程に含む。

本研究のロジスティック回帰分析と共通する考え方は Zajonc (1968) も繰り返し述べているが、具体的な予測式は示していない。今後、自然言語で観察される単純接触効果の研究を進めるにあたり、選好の心的プロセスを視野に入れた計量的予測モデルを手にしておくことはきわめて重要である。なお、本研究では選好 (preference) と選択 (choice) を区別しないで用いることがある。

2. 単純接触効果に関連した価値関数の設定

Kahneman & Tversky (1979) は投資家の選択行動を「価値関数 (Value Function)」で説明する理論を構築し、Kahneman は 2002 年のノーベル経済学賞を受賞した (Tversky は 1993 年に没)。価値関数とは、利得や損失によってもたらされる心理的な価値評価を示す関数である。この発想は、字体選択行動の研究にも有効であろう。以下、経済学とは違う新たな視点から価値関数について検討した。

まず、字体選好はその字体を選択した場合にもたらされる利得と損失のバランスによって左右されるという仮定を設けた。利得と損失の収支（トレードオフ）をうまく捉えられる価値関数が見つかれば、新字体と旧字体のいずれを選ぶかは価値関数で予測することが可能になる。さらに、人間は価値関数を最大化する方向に動くという仮定を付け加えるならば、経済学など言語研究以外への理論的寄与も期待できる。

次に、価値関数に関係がありそうな要因として、字体と人間との親疎関係に着目した。先に示した図 1 の流れをさらに細分化して「親疎関係 → 接触頻度 → 親近度の変化 → 単純接触効果 → 好意度の変化」というプロセスを想定すれば、親疎関係は結果的にその字体に対する好意度を左右すると考えられる。ある対象に抱く好意度の高さは、その対象を選択した（得た）場合の利得だけではなく、逆に、その対象を選択しなかった（失った）場合の損失にも深く影響すると言えよう。

ただし、字体と人間との親疎関係や、字体に対して人間が抱く好意度を、直接的かつ網羅的に測定したデータはいまだ存在しない。言語生活における文字刺激と人間との関係を精度よく調査する方法論が十分には成熟していないのがその原因である。そこで、本研究では、親疎関係や好意度をうまく捉えるための近似的な指標として、親近度を挙げた。先述の通り、親近度は親疎関係や好意度と密接な関係にあることは間違いない。ある字体と人間との親疎関係が変化すると、接触頻度も変化し、結果的に親近度も変化する。また、親近度は、JIS X0208-1983 の第 1・第 2 水準に含まれる漢字 6,355 字種について、天野・近藤（2000）によって測定されたデータが利用可能である。さらに、親近度は頻度との関連性も詳しく分かっているので数理モデルの展開に都合がよい。

以上の議論をふまえた価値関数 V を(1)式に示す。旧字体親近度（familiarity of traditional form）を $Fam1$ 、新字体親近度（familiarity of simplified form）を $Fam2$ とし、両者の差が利得と損失の収支になると仮定した【末尾注 1】。これは、旧字体を選択したことによる利得が $Fam1$ で、新字体を選択しなかったことによる損失が $Fam2$ であることを意味する。なお、(1)式の見た目は親近度の収支計算であるが、その本質は単純接触効果がもたらす好意度の収支計算を近似的に織り込んでいる点を再度強調しておく。単純接触効果を念頭においた価値関数の提唱は、経済学や心理学にも前例がない。

$$V = Fam1 - Fam2 \quad \text{----- (1)}$$

次に, Fam1, Fam2 の項を検討した。

3. 親近度と接触頻度の関係

フェヒナーの法則は, 精神物理学の分野では基本的な法則としてよく知られている。刺激から生じる感覚尺度を Y , 刺激の強度を I , 対数関数を \log (底は e), 傾きを K (定数), Y 軸の切片を C (定数) とする次の単純な関数で表現される。

$$Y = K \times \log(I) + C \quad \text{----- (2)}$$

これを漢字親近度 (なじみ) と漢字頻度の関係に適用すると

$$\text{漢字親近度} = K \times \log(\text{漢字頻度}) + C \quad \text{----- (2.1)}$$

先に述べたように, 新聞における文字の出現頻度データは, それに人間が接触する確率を代表する指標だと仮定して, 接触頻度と親近度との関係を次のように定式化する。旧字体親近度と旧字体頻度 (frequency of traditional form) の関係は, 旧字体親近度 $Fam1$, 旧字体頻度 $r1$ として, 傾き K , 切片 C の関数を考えると【末尾注2】

$$Fam1 = K \times \log(r1) + C \quad \text{----- (2.2)}$$

同じく新字体親近度と新字体頻度 (frequency of simplified form) の関係は, 新字体親近度 $Fam2$, 新字体頻度 $r2$ として

$$Fam2 = K \times \log(r2) + C \quad \text{----- (2.3)}$$

親近度と頻度に関するこのような定式化は, 心理学の分野では珍しいものではない。なお, 漢字頻度が 0 の場合は自然対数の計算ができないため, 予測式で $r1, r2$ の項は, 回帰分析等においては $r1+1, r2+1$ に変換して計算を行う。この変換法は多くの先行研究でしばしば用いられてきたものである。

さて, (1)式の価値関数 V の $Fam1$ に(2.2)式を, $Fam2$ に(2.3)を代入して整理すると

$$\begin{aligned}
 V &= K \times \{ \log(r_1) - \log(r_2) \} \\
 &= K \times \log(r_1/r_2) \quad \text{----- (2.4)}
 \end{aligned}$$

以上により、単純接触効果を織り込んだ価値関数 V は、旧字体頻度と新字体頻度の比で決まるという仮説が導かれた。横山 (2006b) は右辺の $\log(r_1/r_2)$ を「接触相対性 (exposure relativity)」と呼んでいる。

次に、この仮説が実証データをうまく予測・説明できるか検証するために、ロジスティック回帰分析を導入するのだが、その前に、オッズとロジットについて説明しておく。

4. オッズとロジット

「桧一檜」などの異体字選択データにおいて、 R_1 を旧字体選択人数、 R_2 を新字体選択人数とする。2肢強制選択課題において選択肢「1 と 2」の「1」を選ぶ確率を p_1 とおくと、「2」を選ぶ確率は $1-p_1$ となる。 $p_1/(1-p_1)$ を「オッズ (odds)」という。 $R_1+R_2=N$ とすると、異体字「1」の選択率は R_1/N で、オッズは(3)式になる。

$$\begin{aligned}
 p_1/(1-p_1) &= p_1/p_2 \\
 &= (R_1/N)/(R_2/N) \\
 &= R_1/R_2 \quad \text{----- (3)}
 \end{aligned}$$

オッズの対数は「ロジット (logit)」あるいは対数オッズという。よって $\log(R_1/R_2)$ はいわば反応ロジットである。同様に(2.4)式の $\log(r_1/r_2)$ は刺激ロジットである。

5. ロジスティック回帰分析の導入

ここでロジスティック回帰分析を導入する。この分析手法は、医学統計やブランド選択研究の分野などで近年きわめて盛んに利用されるようになった。アスベスト (石綿) などのリスクファクターに曝露 (exposure) されたケースとそうでないケースでは肺疾患の発症率がどの程度高まるのかを推定する場合や、手術から5年後の生存率にいかなる要因が影響を与えるのかなどを探るのに有効だとされている。目的変数が「発症する—しない、

生存-死亡」のように2値をとるデータに適用するのが一般的である。

旧字体選択確率を p_1 とし、それを新字体選択確率で除した値 $p_1/(1-p_1)$ は旧字体選択確率のオッズになる。その自然対数を求めた値すなわち反応ロジットを目的変数とし、説明変数を(1)式の価値関数とするロジスティック回帰分析は、傾き w 、定数 b^* として

$$\log \{ p_1 / (1 - p_1) \} = w \times V + b^* \quad \text{----- (4)}$$

左辺に(3)式を、右辺の V 項に(2.4)式を代入し、 $w \times K$ を a とおくと

$$\begin{aligned} \log (R_1 / R_2) &= w \times K \times \log (r_1 / r_2) + b^* \\ &= a \times \log (r_1 / r_2) + b^* \quad \text{----- (4.1)} \end{aligned}$$

これを変形すると

$$R_1 / R_2 = \exp(b^*) \times (r_1 / r_2)^a \quad \text{----- (4.2)}$$

次に、このモデルと動物行動研究の関係について検討した。

6. 一般対応法則との関係

一般対応法則 (generalized matching law) は動物の選択行動研究から Baum (1974) が導いたもので、「反応比」(R_1/R_2) と「強化比」(r_1/r_2) が単純な関数関係のもとで対応していることをいう。反応比、強化比とは、次のような意味である。ハトやラットなどの動物がキーを押すと餌(報酬)がもらえる環境を実験室内に準備し、2つのキーを設定する。例えば、右側のキーを「1」、左側のキーを「2」としよう。「1」を押すと報酬が出る頻度は r_1 、「2」を押すと報酬が出る頻度は r_2 である。この報酬頻度の比を強化比という。訓練(学習)を受けた動物のキー押し反応を観察すると、「1」を押す反応頻度は R_1 、「2」を押す反応頻度は R_2 となる。この反応頻度の比を反応比という。

左右キーの反応比と強化比は、次の(5)式の形で対応することがハトやラットやマウスなどによる数多くの研究で示されている (Woolverton & Alling, 1999; Belke & Belliveau, 2001; 乙部, 2005)。

$$R1/R2 = b \times (r1/r2)^s \quad \text{----- (5)}$$

bは反応の偏り（バイアス）を、sは反応の敏感さ（感度）を意味するパラメータである。
 $s=a$, $b=\exp(b^*)$ とおけば、(5)式は(4.2)式と一致する。

また、(5)式の両辺の自然対数をとって展開すると

$$\begin{aligned} \log(R1/R2) &= \log \{ (r1/r2)^s \times b \} \\ &= s \times \log(r1/r2) + \log(b) \quad \text{----- (5.1)} \end{aligned}$$

(5.1)式は目的変数が $\log(R1/R2)$ 、説明変数が $\log(r1/r2)$ の回帰式である。つまり、反応ロジットを刺激ロジットから予測する式になっている。 $s=a$, $\log(b)=b^*$ とおくと、(5.1)式は(4.1)式と一致する。よって、単純接触効果を織り込んだロジスティック回帰分析は一般対応法則と数式のうえで同じになることが証明された。(4.1)式は人間の精神物理学に基づいて導出されたものであるが、分野の異なる動物行動学の知見と矛盾しない点は興味深い。

7. 接触相対性理論の提唱

異体字選択を予測する要点は、(2.4)式や(4.2)式などが示すように、新旧字体の頻度によるロジット $\log(r1/r2)$ の値にある。以下、横山(2006b)にしたがって $\log(r1/r2)$ を接触相対性という。また、単純接触効果から接触相対性を導く一連の考え方を「接触相対性理論(exposure relativity theory)」という。この予想がどの程度までの中するかを、字体選好課題の実験データによって計量的に検証した。

8. 字体選好実験によるモデル評価

接触相対性理論の妥当性について、(5.1)式の予測力を中心に検証を行った。

実 験

方 法

刺激材料 刺激項目は新旧 263 ペア（笹原・横山・ロング, 2003）。JIS X0208-1983 の第

1・第2水準に含まれる漢字で、新字体（拡張新字体）と旧字体（正字体）の関係にあるもの。処理が複雑になるため、JIS 漢字に含まれる異体字の中でほとんど使われないものは原則として扱わなかった。異体字ペアの例を図2に示す。

01	亜	亞	09	葛	葛
	啞	啞		喝	喝
	壺	壺			
02	媛	媛	10	觀	觀
	淫	淫		灌	灌
	稗	稗	11	爛	爛
03	陷	陷		澗	澗
	焰	焰	12	徽	徽
04	奥	奧	13	俠	俠
	襖	襖		狹	狹
				頰	頰

図2 異体字ペアの例

漢字頻度データ 263ペアのうち新旧字体ともにJIS X0208-1983で表示可能なペア86組を抽出した。これ以外の177ペアは旧字体がJIS X0208-1983で表示できないこともあり、旧字体の頻度データについて信頼に足る資料が存在しないため、以後の分析からは除外した。異体字ペア86組の新旧字体頻度は、朝日新聞データ（横山・笹原・野崎・ロング, 1998）を用いた。

手続き 実験の冒頭で「ワープロを打っている場面だけをイメージするように」と伝え、異体字のペアを実験参加者に呈示して、より使いたいと感じる方の字を選択させた。具体的な教示は次の通り。「この実験は、漢字の使われ方を調べるものです。これから、字の形は違いますが、読みと意味がまったく同じ漢字のペアをお見せします。たとえば「断」と「斷」は、同じ読みで同じ意味の漢字のペアです。もし、あなたがワープロを打っているとしたら、どちらの字を使いたい、教えてください。2つの漢字をよく見て、使いたいと感じる程度を比較し、より使いたいと思う方の字に○印をつけてください。両方とも使いたい、あるいは両方とも使いたくないと感じるペアがあるかも知れませんが、とにかく、どちらか一方の字だけに○印をつけてください。判断は、あまり深刻に悩まずに、直観的

に行ってください。(以下略)」

実験参加者 京都市内の立命館大学の学生 72 名（男性 20 名，女性 52 名）。このデータは横山（2006b）が分析に使用したものとは異なる。

結果と考察

（1）ロジスティック回帰分析の解析結果

旧字体頻度を r_1 ，新字体頻度を r_2 とする接触相対性 $\log(r_1/r_2)$ と，旧字体選択率を表 1 に示す。接触相対性が小さくなると旧字体選択率も低下する傾向が見える。

表 1 字体ペアの接触相対性で排列した旧字体選択率の一部

ID	新旧 字体ペア	接触相対性 $\log(r_1/r_2)$	旧字体 選択率 (%)	ID	新旧 字体ペア	接触相対性 $\log(r_1/r_2)$	旧字体 選択率 (%)
68	尔爾	3.72970	62.5	179	獵獵	-5.2417	6.9
138	鈇鐸	3.66356	58.3	108	尽盡	-5.9713	2.8
39	頸頸	1.92181	83.3	164	万萬	-5.9945	11.1
69	迤邐	1.79176	45.8	170	謡謡	-6.0186	11.1
182	蕊蕊	1.60944	33.3	78	肅肅	-6.2860	8.3
176	籠籠	1.30716	79.2	84	将将	-6.2932	2.8
122	賤賤	1.20397	72.2	152	麦麥	-6.4184	0
177	鼠鼠	0.09531	83.3	85	獎獎	-6.6120	9.7
		(中略)				(中略)	
125	曾曾	-4.1929	27.8	110	数數	-9.1915	0
104	楨楨	-4.3438	40.3	172	来來	-9.6622	4.2
100	真真	-4.4026	4.2	58	参参	-9.7406	0
93	条條	-4.4431	9.7	17	会會	-9.9100	2.8
171	瑶瑶	-4.6770	33.3	148	壳賣	-10.063	0
95	嬢嬢	-4.6913	16.7	34	区區	-10.254	0
50	広廣	-4.7780	9.7	155	発發	-10.264	0
1	亜亞	-5.1514	1.4	140	戦戦	-10.536	1.4

新旧字体選択比（選択人数比：オッズ）から $\log(R_1/R_2)$ を求める際に， R_1 か R_2 がゼロになると対数値が無限大になる。それを避けるために「経験ロジット」に変換した (Collett, 2003) うえで(5.1)式の s と $\log(b)$ を推定したのが(6)式である【末尾注 3】。

$$\log(R_1/R_2) = 0.41 \times \log(r_1/r_2) - 0.25 \quad \text{----- (6)}$$

接触相対性 $\log(r1/r2)$ と目的変数 $\log(R1/R2)$ の相関図を図 3 (1) に示す。接触相対性と目的変数の積率相関は $r=.81$ ($p<.01$, $df=84$)、説明率は 65.51% であった。

ちなみに、接触相対性 $\log(r1/r2)$ と旧字体選好率の相関図を図 3 (2) に示す。積率相関は $r=.77$ ($p<.01$, $df=84$)、説明率は 59.40% であった。線形性がかなり崩れていることから、目的変数は(6)式のようにロジットの方が望ましいと言える。

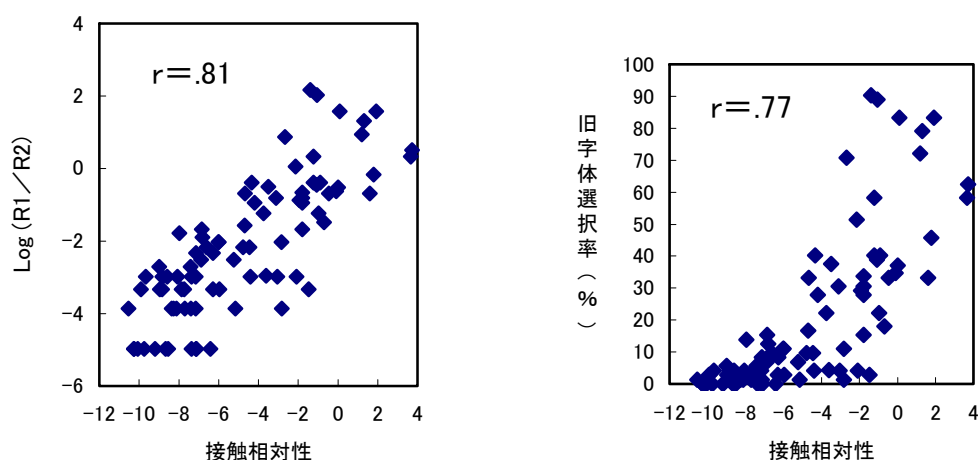


図 3 (1) 接触相対性と $\log(R1/R2)$ の相関図

図 3 (2) 接触相対性と選好率の相関図

(2) 親近度と字体選好率の関係

漢字親近度データ (天野・近藤, 2000) を(1)式の評価関数 $V = \text{Fam1} - \text{Fam2}$ に代入して求めた値と、旧字体選好率 (%) との積率相関は $r=.89$ ($p<.01$, $df=84$) であった。この数値は(6)式による説明変数と目的変数の相関 $r=.81$ より高い。これは評価関数 V の妥当性をうかがわせるものだと考えられる。

ある文字の使用頻度が高い場合は、人間がその文字に接触する確率が高くなり、接触確率が高くなると親近性が増加するとともに単純接触効果が生じて選好度も高くなると考えられる。つまり、「使用頻度 → 接触確率 → 親近度 → 選好度」となり、頻度よりも親近度の方が選好度に直接影響する要因だと言えよう。

(3) 字体間での形態差の大きさと選好率の関係

字体間に微細な差しかないペアについての選好率を分析した。例えば、「奥—奥, 讚—讚, 騷—騷」は一見どこに違いがあるかの識別が困難で、視覚的な混同 (confusion) が生じる

確率が高いため、新旧字体がランダム選択されて旧字体選好率はチャンスレベルの50%前後になる可能性もあるだろう。しかし、この3ペアのそれぞれについて新旧字体の選好人数をカイ2乗検定した結果、新字体を選好する人が有意に多いことが示された（それぞれ $p < .01$, $df = 1$ ）。選好率は、「奥 97.2% : 奥 2.8%」、「讚 95.8% : 讚 4.2%」、「騷 98.6% : 騷 1.4%」であった。その他の260ペアについても同様にカイ2乗検定を実施したところ、字体の違いが小さい場合に新旧字体で選好率がチャンスレベルの50%ずつになるというような傾向は確認されなかった。その典型例を図4に示す。これは日本語学で「筆法差」といわれるものである（笹原・横山, 1998）。

挽 挽

図4 字体間に微細な差しか存在しない例

このように字体間におけるデザイン差の大小が視覚的混同に影響しなかった理由として、刺激ペアを瞬間呈示したのではなく実験参加者ペースで課題を実施したことが考えられる。実験参加者に与えられた教示からペア間に何らかの形状差やデザイン差があることは自明であり、その示差的特徴（*distinctive feature*）を慎重に検出したうえで選好判断を行うという「心的構え」が実験参加者にできていたのであろう。このような処理の構えが、日常の言語生活でいつも用いられているかどうかは今後の検討を待つ必要がある。

（4）文字構成要素と選好率の関係

文字構成要素を共有する漢字グループをピックアップして選好率を見ると、「会 97.2% : 會 2.8%」で新字体が有利なのに、木偏がつくと「桧 19.2% : 檜 70.8%」というように旧字体有利に逆転するというケースがあった。同様の現象は、「観 97.2% : 觀 2.8%」と「漚 9.7% : 灌 90.3%」のほか、「経 98.6% : 經 1.4%」と「頸 16.7% : 頸 83.3%」に見られた。この結果から、人間は字体選好判断を漢字の構成要素に分解して行うのではなく、漢字全体をゲシュタルトとしてとらえて判断する場合もあると考えられる（笹原・横山, 1998）。

9. まとめ

異体字選択における意思決定が漢字頻度からどのくらい予測できるかについて検討した。異体字とは「桧—檜」のように読みと意味は同じで字体だけが異なる漢字の集合を指す。

社会心理学の単純接触効果，経済学の価値関数，精神物理学のフェヒナー法則，動物行動学の一般対応法則などをロジスティック回帰分析に結びつけ，新聞漢字頻度から旧字体選択率を予測する接触相対性理論を導出した。この理論の妥当性を検証するため，字体選好課題のデータを収集して予測力を評価した。字体選好課題とは，情報機器で漢字を書いている場面を実験参加者にイメージさせ，異体字 263 ペアについての選好判断を 2 肢強制選択法で求めるものであった。その結果，漢字頻度に基づいて算出した接触相対性と実際の観測値の間には強い正の相関関係がみられ ($r=.81$)，予測式の説明率は 65% に達した。これから，異体字選好は新旧字体ペア間での接触頻度の相対比によって 6 割以上が決定されることが明らかになった。人間と字体との接触頻度が増加すると単純接触効果が生じ，その字体に対する好意度が高くなる，という流れによって字体の選択率が変化すると考えられる。

本研究は，ある仮定のもとで，接触相対性を説明変数とするロジスティック回帰分析が一般対応法則と同じになることを証明した【末尾注 4】。以下に(5.1)式を再掲する。

$$\log (R1/R2)=s \times \log (r1/r2)+\log (b) \quad \text{----- (5.1)}$$

上記の「ある仮定」とは，選択肢間での親近度の差を，接触頻度の違いから求めるために，フェヒナー法則を組み込んだ価値関数を設定したことを指す。(5.1)式は，いわば「接触効果方程式 (exposure effect equation)」になっており，単純接触効果，一般化対応法則，フェヒナー法則の 3 者は密接に関連していると考えてよい。以上の事実は，人間，ハト，ラットなどの種の違いを越えた，動物に普遍的な選択・選好メカニズムが存在する可能性を示唆している。

この接触方程式の妥当性や適用可能性などについては今後の検討を待つ必要があるが，経済心理学分野での竹村・藤井 (2005) による一般対応法則とランダム効用理論についての議論は示唆に富む。彼らは，ランダム効用理論とフェヒナー法則の両者を用いて，ランダム効用最大化過程の誤差項にガンベル分布を仮定して積分計算を行い，一般対応法則を導出することに成功している。竹村・藤井 (2005) と本研究による(5.1)式の導出方法は，一脈通じるところがあると言えよう。

さらに，野生の動物集団がいくつかの餌場間でどのように分布するかを説明する「理想自由分布理論 (the ideal free distribution theory)」について検討した山口・伊藤 (2006)，Fagen

(1987) などのモデル式も(5.1)式と一致する。ロジスティック回帰分析の適用範囲はかなり広いのではないかとの期待を抱かせる。

Zajonc は単純接触効果の研究を始めた 1968 年以降、一貫して漢字刺激を使い続けてきた。その理由は、おそらく、漢字刺激を用いると実験がうまくいくからではないかと推察する。漢字は視覚パターンとして凝集性があり、偏や旁といった構造を有している点でランダム図形パターンとは一線を画する。非漢字圏実験参加者にとって、漢字は新奇刺激であるが、何回か接触するうちに経験の痕跡が心内に蓄積されやすいという性質を持っているのかもしれない。誰も初めて見る漢字は新奇刺激であるが、学校教育や日常生活で意識的・無意識的にそれを目にしているうちに、その漢字刺激に対する好意度が高まっていくのであろう。漢字のほかに、図形、写真、広告ロゴなどを刺激とした場合の単純接触効果の研究は生駒 (2005) に詳しい。

また、年代差などの社会言語学的な眼を心理学に導入した研究も必要とされている。日本の方言研究が世界をリードしてきたのは、国立国語研究所などが中心になって話し言葉の変異を地域差 (全国約 3,000 地点) と年代差 (10 歳代から 60 歳代以上) にわけて丹念に精査してきたことによる。言語の変異は方言だけに観察されるものではなく、異体字選好課題にも見られる。日本全国から調査参加者を無作為抽出し、Web 画面で「桧—檜」の異体字ペアを呈示して調査を行ったところ、年代差が確認されたとの報告がある (横山, 2004)。調査対象は 20 歳代、30 歳代、40 歳代、50 歳代の 4 群で、各群 120 名ずつ計 480 名であった。調査の教示は本研究の字体選好課題とほぼ同じであった。その結果、20 歳代は旧字体「檜」を選好する割合が 7 割近くを占めて新字体「桧」の 3 割を上回ることが示された。ところが、40 歳代になると旧字体「檜」が約 4 割に減少し、逆に新字体「桧」を選択した人が約 6 割に増加して逆転現象が生じた。さらに 50 歳代では「檜」が約 3 割に落ち込み、「桧」が約 7 割に達した。つまり、20 歳代と 50 歳代は異体字の選択傾向はまるで逆であり、中年層以上は旧字体「檜」をあまり選択しないことが明らかになった。おそらく、若者は手で書くことを意識していないのに対して、中年層は手で書く場合の筆記の経済性を何らかの形で考慮に入れているのだろう。あるいは、老眼の影響があって、複雑な字形は読みにくいと感じるのかもしれない【末尾注 5】。Web 調査はサンプルが偏る危険性もあるので一般的な結論は差し控えなければならないが、異体字選好課題には年代要因が影響する場合もあると予想できる。

そのほか、文字よりも大きな構成単位である語や文のなかで異体字がどのように使われ

るのか、という問題も検討すべきであろう。語や文が異体字選好に対してある種の文脈を与える可能性がある。仮にそうならば、語や文の文脈による影響と、漢字の親近度による影響が、どのような関係を結ぶのかは非常に興味深い研究テーマになるだろう【末尾注 6】。今後はこのような側面にも注意を向けながら研究を進める必要がある。

【注 1】 価値関数は利得と損失の収支を心の中で計算するものであり、さまざまな計算法を想定できるが、本研究では一番単純な演算として親近度差を採用した。

【注 2】 (2.2)式と(2.3)式の定数 K , C は、すべての文字でほぼ同一であり、違いを無視しても差し支えないという前提を立てた。文字ごとに、あるいは異体字ペアごとに定数 K , C が異なるモデルを想定すると、(2.4)式の価値関数 V がきわめて複雑な形になり、推定すべきパラメータが爆発的に増えてしまう。パラメータの数を最小限に抑えるために、大過ないと考えられる前提を採用した。

【注 3】 オッズ($R1/R2$) が 0 もしくは 1 になって対数計算が不能になる問題を回避するため、選択率を $p^*=(r+0.5)/(n+1)$ に変換してオッズ $p^*/(1-p^*)$ の自然対数を求める「経験ロジット (empirical logit)」が慣習的に用いられる (Collett, 2003)。本研究はこれに従った。また、ロジスティック回帰分析のパラメータ推定は「最尤法」で求める場合が多いが、ここでは最小 2 乗法により求めた。最小 2 乗法を採用した理由は、実用上の近似解として差し支えない精度を達成すると期待できることと、最尤法の計算に必要なソフトウェアを必要としない点を考慮したことによる。

【注 4】 本研究は 2 肢選択 (1 対比較) のデータを扱ったが、選択肢が 3 つ以上ある場合は「多項ロジスティック回帰分析」を適用すればよい。

【注 5】 野山広 (2006, 私信) による。

【注 6】 この観点は査読者から授かった。

謝辞

本研究で使用した字体選好実験のデータは、當山日出夫氏 (花園大学)、笹原宏之氏 (早稲田大学社会科学部) と横山が 1998 年 4 月に立命館大学 (京都市) で採取したものである。データ使用の許諾をくださった當山氏と笹原氏に深く感謝申し上げる。また、英文要旨はエリック＝ロング氏に校閲をしていただいた。ロング氏にも感謝の意を表す。

引用文献 (アルファベット順)

- 天野成昭・近藤公久 [編著] (2000) 『日本語の語彙特性』NTT データベースシリーズ, NTT コミュニケーション科学基礎研究所 [監修], 三省堂
- Baum, W.M. (1974) On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior*, **22**, pp.231-242
- Belke TW, & Belliveau J. (2001) The general matching law describes choice on concurrent variable-interval schedules of wheel-running reinforcement. *Journal of the Experimental Analysis of Behavior*,

- 75(3), pp.299-310.
- Collett, D. (2003) *Modelling Binary Data*. Chapman and Hall, London, United Kingdom.
- Fagen, R. (1987) A generalized habitat matching rule. *Evolutionary Ecology*, **1**, pp.5-10
- Herrnstein, R. J. (1961) Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, **4**, pp.267-272
- 生駒 忍 (2005) 「潜在記憶現象としての単純接触効果」『認知心理学研究』3 巻 1 号 pp.113-131, 日本認知心理学会
- Kahneman, D., & Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, pp.263-291
- 乙部貴幸 (2005) 「近交系マウスの報酬遅延に対する感度」『日本心理学会第 69 回大会発表論文集』 p.1073
- 笹原宏之・横山詔一 (1998) 「異体字選択に影響する要因」『計量国語学』21 巻 7 号 pp.291-310, 計量国語学会
- 笹原宏之・横山詔一 (2000) 「異体字に対するなじみと好み——接触印象・使用頻度との関係」『日本語科学』8 号 pp.110-125, 国立国語研究所〔編〕, 国書刊行会
- 笹原宏之・横山詔一・エリック＝ロング〔著〕 (2003) 『現代日本の異体字——漢字環境学序説』国立国語研究所プロジェクト選書No.2, 三省堂
- 竹村和久・藤井 聡 (2005) 「一般化マッチング法則とランダム効用理論」『日本心理学会第 69 回大会発表論文集』 p.807
- Thorndike, E. L., & Lorge, I. (1944) *The teacher's word-book of 30,000 words*. Teacher's College, Columbia University, New York
- Woolverton WL, & Alling K. (1999) Choice under concurrent VI schedules: comparison of behavior maintained by cocaine or food. *Psychopharmacology*, **141**(1), pp.47-56.
- 山口哲生・伊藤正人 (2006) 「理想自由分布理論に基づく個体分布の実験的検討——絶対報酬量と集団サイズの効果」『心理学研究』76 巻 6 号 pp.547-553
- 横山詔一 (2004) 「文字処理の認知科学」, 『言語』, 33 巻 8 号 pp.56-63, 大修館書店
- 横山詔一 (2006a) 「潜在記憶と言語習得」, 『言語』, 35 巻 4 号 pp.52-57, 大修館書店
- 横山詔一 (2006b) 「異体字選好は新聞漢字頻度から予測可能か」『計量国語学』25 巻 4 号 pp.181-194, 計量国語学会
- 横山詔一・笹原宏之 (2000) 「文字と暮らし」『豊かな言語生活のために』(新「ことば」シリーズ 11) pp.52-63, 国立国語研究所〔編〕, 大蔵省印刷局
- 横山詔一・笹原宏之・野崎浩成・エリック＝ロング〔編著〕 (1998) 『新聞電子メディアの漢字——朝日新聞 CD-ROM による漢字頻度表』国立国語研究所プロジェクト選書No.1, 三省堂
- Zajonc, R.B. (1968) Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, **9**, pp.1-27

(2006 年 3 月 14 日受付)

(日本語抄録)

論文

字体選好は新聞漢字頻度から予測可能か

横山詔一 (国立国語研究所)

異体字選択における意思決定が漢字頻度からどの程度予測できるかについて検討した。異体字とは「桧一檜」のように読みと意味は同じで字体だけが異なる漢字の集合を指す。社会心理学の単純接触効果、経済学の価値関数、精神物理学のフェヒナー法則、動物行動学の一般対応法則、生態学の理想自由分布理論などをロジスティック回帰分析に結びつけ、新聞漢字頻度から旧字体選択率を予測する接触相対性理論を導出した。この理論の妥当性を検証するため、字体選好課題のデータを収集して予測力を評価した。字体選好課題とは、情報機器で漢字を書いている場面を実験参加者にイメージさせ、異体字 263 ペアについての選好判断を 2 肢強制選択法で求めるものであった。その結果、漢字頻度に基づいて算出した接触相対性と実際の観測値の間には強い正の相関関係がみられ ($r=.81$)、予測式の説明率は 65% に達した。これから、異体字選好は新旧字体ペア間での接触頻度の相対比によって 6 割以上が決定されることが明らかになった。人間と字体との接触頻度が増加すると単純接触効果が生じ、その字体に対する好意度が高くなる、という流れによって字体の選択率が変化すると考えられる。

(英文アブストラクト)

PAPER

Mere Exposure Effect and Generalized Matching Law on Preference for Kanji Form.

YOKOYAMA Shoichi (National Institute for Japanese Language)

Descriptors: preference, mere exposure effect, value function, matching law, Fechner's law, familiarity

Abstract :

In writing the word *hinoki* 'cypress' in kanji, there are two commonly-used alternative forms, 檜 (the so-called "traditional" form) and 桧 (a "simplified" form), both of which have the same meaning and pronunciation. Such alternative forms are called kanji variants. The study introduces a logistic regression analysis, based on "Fechner's law", which predicts familiarity judgments performance from frequency of kanji variant. This model was verified on empirical data obtained from an experiment, in which 72 subjects in Kyoto were presented 263 traditional/simplified variant pairs and asked which form they would prefer to use in word processing. Logistic regression analysis were conducted with the logit for choice of the traditional form as the dependent variable, and the logit of the frequency of the characters as the independent variable (cf. Formula 5.1) and results showed significant correlations($r = .81$ for Asahi). This logistic regression analysis is very similar to "generalized matching law" by Baum (1974) or "the ideal free distribution theory" by Fagen(1987).