

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)

ディスクリプタ : 異体字, 選好, 親近度, フェヒナーの法則,
単純接触効果, 2 肢強制選択法

0. はじめに

漢字には「桧一檜」のような異体字 (variant) が少なからず存在する。笹原・横山 (1998) は, 異体字研究を漢字研究の旧来の枠組みのなかだけで行うのではなく, 社会言語学や認知科学も射程に入れて, 有効な調査・実験手法を編み出すことを目標に掲げ, 「選好 (preference : 好み)」や「親近度 (familiarity : なじみ)」といった概念を経済学や心理学から導入して一連の研究を行ってきた。その手法は簡単明瞭で, 単漢字の異体字ペアを被験者に呈示して「字体選好課題」あるいは「字体親近度比較課題」を実施するというものであった。字体選好課題は, 263 字種の異体字 (新旧字体) ペアを呈示し, パソコンや携帯メールなど情報通信機器 (IT 機器) で字を書いているとしたらどちらの字体を使いたいかを被験者に 2 肢強制選択法 (2-Alternatives Forced Choice) で直観的に選ばせるという課題である。字体親近度比較課題は, 新旧字体ペアのうち, 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。東京都内の女子大学生約 200 名を対象にこの手法で収集された素データは電子媒体で公開されている (笹原・横山, 2000)。

本研究は, この公開データに基づき新旧字体ペア 263 字種すべてについて各ペアの旧字体選好率を算出し, その値が朝日新聞ならびに読売新聞の漢字頻度からどのくらい予測可能かを検討した。新聞の漢字頻度データなどから字体選好課題の結果をある程度正確に予測できるモデルが手に入れば, 文字生活に関する新たな研究パラダイムを開く契機にもな

Can We Predict Preference for Kanji Form from Newspaper Data on Character Frequency?

YOKOYAMA Shoichi (National Institute for Japanese Language)

ると考えられる。文化庁は、文字表記や会話における言い回しに関する調査を経年的に実施している。例えば、『国語に関する世論調査』（文化庁、2005）では「あいさつ—挨拶」や「ひんしゆく—輦蹙」などでどちらをより好むかという表記の選好に関する質問項目を含んでいる。言語生活はどの言語表現を選択するかという判断の連続であり、いわば言語表現の選好によって支えられている部分が多い。その背後に潜む要因やメカニズムを綿密に検証する際には、計量的な理論やモデルが威力を発揮するであろう。

さらに、人間心理における選好のメカニズムは、記憶や認知と深く関係していることが最近の研究で明らかになっている。社会心理学の分野で Zajonc (1968) が発見した単純接触効果 (mere exposure effect) は、新奇刺激に繰り返し接触するだけで、その刺激に対する選好度が高まるという現象である。この効果は意識下でも生じることが世界各国の多くの研究で実証されており、信頼性が高い。その実験はいろいろな方法があるが、脳科学研究の分野では次のような研究例がある。英国人 9 名の被験者に対して、彼らが見たことのない漢字 20 字 (隊, 謙, 働など) を 1 文字ずつ 0.05 秒間だけ瞬間呈示し、続けてマスクング刺激を 0.45 秒間呈示した。各漢字は 10 回ずつ呈示され、刺激呈示に要した時間は合計で 1 分 40 秒であった。このような条件下で英国人が新奇な漢字刺激を知覚するのはほぼ不可能であり、その漢字刺激を見たという接触意識さえ持てない。ところが、刺激呈示終了後に、経験済みの漢字とそうでない漢字をペアで示して、どちらをより好むか 2 肢強制選択法で尋ねると、経験済みの漢字が選ばれる傾向にあったと報告されている (Elliot & Dolan, 1998)。

この単純接触効果は、新奇刺激を意識的・無意識的に経験するにつれて、刺激に対する被験者の親近度が高まるために生じる、という説が有力な考え方の一つとして知られている (Monin, 2003)。これらの知見をふまえて、本研究では字体選好の基本メカニズムとして単純接触効果によって生じる個々の漢字に対する親近度 (なじみ度) の増加を仮定した。その妥当性を吟味するために、漢字 1 文字の親近性から字体選好課題の結果がどのくらい正確に予測できるのかという問題についても検討を行った。

1. 精神物理学に基づく予測モデルの導出

1. 1. 漢字頻度から親近度を予測するモデル

本研究では、精神物理学の感覚知覚分野でよく知られているフェヒナーの法則 (Fechner's law) に着目し、朝日新聞や読売新聞の漢字頻度データから字体選好課題の結果を予測する

モデルを導出することにした。

フェヒナーの法則は、ある刺激から生じる感覚尺度を S 、刺激の強度を I 、常用対数を Log 、傾きを K (定数)、 S 軸の切片を C (定数) とする次の線型結合で表現される。

$$S = K \times \text{Log}(I) + C \quad \text{-----}(1)$$

(1) 式から漢字親近度と漢字頻度の関係について類推をはたらかせると

$$\text{漢字親近度} = K \times \text{Log}(\text{漢字頻度}) + C \quad \text{-----}(2)$$

新旧字体ペアの旧字体刺激から生じる親近度 (familiarity of traditional form) を「旧字体親近度」、その旧字体が新聞に出現した頻度 (frequency of traditional form) を「旧字体頻度」として、傾き K 、切片 C の線型結合を考えると次のようになる。

$$\text{旧字体親近度} = K \times \text{Log}(\text{旧字体頻度}) + C \quad \text{-----}(3)$$

同じく「新字体親近度」(familiarity of simplified form) は

$$\text{新字体親近度} = K \times \text{Log}(\text{新字体頻度}) + C \quad \text{-----}(3.a)$$

なお、漢字頻度がゼロの場合は常用対数の計算ができないため、あらかじめ頻度データに 1 を加算しておく。本論文に示す予測式で旧字体頻度、新字体頻度の項は、回帰分析等においては旧字体頻度+1、新字体頻度+1 に変換して計算を行うものとする。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。

1. 2. 漢字頻度から字体選好課題の結果を予測するモデル

ある新旧字体ペアの「旧字体選好率」は字体親近度比較によってなされると仮定し、新旧字体ペア間での親近度差 (旧字体親近度から新字体親近度を減じた値) を変数として、傾き a 、切片 b の線型結合で表現するならば

$$\text{旧字体選好率} = a \times (\text{旧字体親近度} - \text{新字体親近度}) + b \quad \text{-----}(4)$$

旧字体親近度に(3)式を，新字体親近度に(3.a)式を代入して整理すると

$$= a \times K \times \{\text{Log}(\text{旧字体頻度}) - \text{Log}(\text{新字体頻度})\} + b \quad \text{-----}(5)$$

$$= a \times K \times \text{Log}(\text{旧字体頻度} / \text{新字体頻度}) + b \quad \text{-----}(5.a)$$

(5.a) 式が示すのは，旧字体頻度と新字体頻度の比を求め，それを対数変換して説明変数とし，旧字体選好率を目標変数とする回帰式を作成すれば，予測モデルが手に入るということである。これを接触相対性モデル (exposure relativity model) と呼ぶ。

新字体選好率についても (4) 式を次のようにすれば同様に求められる。

$$\text{新字体選好率} = a \times (\text{新字体親近度} - \text{旧字体親近度}) + b \quad \text{-----}(4.a)$$

なお，これらのモデルは新旧字体ペア間での画数差なども加えた重回帰モデルに展開することも可能であるがその検討は別の機会に譲る。以下，おもに (5.a) 式の予測精度について吟味した。

2. 方法

2. 1. 字体選好課題について

すでに CD-ROM で公刊されている笹原・横山 (2000) から，新旧字体ペアの両者が 83JIS 漢字で表示・印刷できる 86 ペアを抽出し，本研究のデータとした。笹原・横山 (2000) の CD-ROM に収録されている 263 ペアのデータ測定方法は以下の通り。

刺激材料：調査に用いた新旧字体ペアの刺激項目は，笹原・横山 (1997) が用いた 181 組の異体字ペアに「蕊－藁」のペア 1 組と，ワープロソフトに現時点ではフォントが実装されていないものの『今昔文字鏡』のフォントに手を加えることで表示・印刷が可能になった異体字ペア 81 組を追加し，全体で計 263 組の異体字ペアを準備した。上記の笹原・横山 (1997) が用いた 181 組の異体字ペアは「桧－檜」「壺－壺」のようなペアで，構成要素によって 81 グループに分類されていた。

それらは以下の 3 つの規準にしたがって選ばれた。①「JIS 漢字」(1997) 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。②「MS 明朝フォント」と「FA 明朝フォント」で各字体が表現できる

もの。この基準を導入した理由は、ワープロソフトによる印字の制約を考慮したことによる。③上記2つの規準に適合した異体字集合から、被調査者になじみがないと思われる字や、字種が多いグループではいくつかの字を原則として削除した。

フェイスシートには年齢、ワープロ使用経験の有無についての回答欄も設けた。調査票の一部を図1に示す。ペアの呈示順序と新旧字体の左右位置はランダム化されていた。

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

図1 字体選好課題に用いた新旧字体ペアの一部

実験の教示：被験者に対する教示は次のようなものであった。いずれの調査においても、場面については、「ワープロを打っている場面だけをイメージするように」とした。ワープロの画面上で文字を選択するのではなく、文字を自らの手で書くのであれば書記経済という点から簡単な字体を選ぶ人が多くなるという可能性もあったが、想定される場面を特定することにより、その影響を小さくするよう心がけた。

字体選好課題の手続き：異体字のペアを被調査者に示して、より使いたいと感じる方の字を選択させた。調査は授業の一環として、集団方式で実施した。調査に先立って、以下の教示を被調査者に与えた。「この調査は、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のか、教えてください。二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)」

被験者：すべて東京都内の女子大学生であった。字体選好課題には 102 名が参加した。被調査者の所属は以下の通り。A 大学文学部国際文化学科・英語英文学科，A 大学短期大学部国際文化学科欧米文化，B 大学学芸学部英文学科・国際関係学科，C 大学文学部日本文学科・英米文学科。これらの被験者には、中国，台湾，韓国，ベトナムからの留学生やワープロの使用経験がない者，あるいは 30 歳前後の者も含まれていた。そこで，25 歳以下で，ワープロの使用経験があると回答した日本人女子学生を抽出した。その結果，85 名のデータを扱うことになった。

2. 2. 新聞の漢字頻度データについて

朝日新聞の漢字頻度は Chikamatsu, Yokoyama, Nozaki, Long, & Fukuda (2000) や横山・笹原・野崎・ロング (1998) によるデータを，読売新聞は文化庁国語課 (2000) のデータを使用した。朝日新聞は 1993 年 1 月 1 日から 12 月 31 日の間に東京本社管内で発行された最終版の朝刊および夕刊で、『CD-HIASK'93 朝日新聞記事データベース』(朝日新聞社，1994) の電子化テキストに基づく。78JIS 漢字と 83JIS 漢字の間でネジレ関係にある異体字群について実際の紙面と比較しながら目視によって確認して頻度を計数した。集計した漢字の延べ数は 17,117,320 (約 1,700 万字) で，漢字の異なり数は 4,583 であった。読売新聞は 1999 年 7 月 1 日から 8 月 31 日の間に東京本社・中部本社管内で発行された最終版の朝刊および夕刊で，テレビ・ラジオ面と広告面は除いたものである。漢字の延べ数は 25,310,226 (約 2,500 万字)，漢字の異なり数は 4,546 であった。

朝日新聞と読売新聞で漢字の延べ数に差があるが，後の分析では頻度を対数変換するために延べ数の違いは軽微な影響しか与えないと見込んで，頻度データを正規化することはしなかった。ちなみに，2500 万の対数値と 1700 万の対数値の差は 0.17 でしかない

2. 3. 漢字の親近度データについて

NTT データベースシリーズ『日本語の語彙特性』(天野・近藤，2000) に掲出されている漢字の親近度(天野・近藤の用語では親密度)のデータを使用した。親近度とは，ある漢字にどの程度のなじみを感じるかの主観的評定値をいう。具体的には，漢字を 1 文字ずつ約 20 名の被験者に視覚呈示し，そのなじみの程度を 1 から 7 までの評定尺度 (1：なじ

みがない----- 7：なじみがある) で評定させて、その平均を求めたものを指す。

3. 分析と結果

先に述べたように笹原・横山(2000)の新旧字体263組からペアの両者が83JIS漢字で表示・印刷できる86ペアを抽出し、旧字体選好率(%)を計算した。その結果を表1に示す。表1には新旧字体ペア間での親近度差(旧字体親近度から新字体親近度を減じた値)と、朝日新聞における新旧字体別の頻度も掲出した。

3. 1. 漢字1文字ごとの親近度データによる予測

漢字1文字ごとの親近度データから、(4)式に基づいて新旧字体ペア間の親近度の差を説明変数、旧字体選好率を目標変数とする回帰分析を行ったところ、表2に示す(4.1)式が得られた。説明変数と目標変数の相関係数は $r=0.86$ ($p<.01$, $df=84$)、説明率は73.34%であった。

3. 2. 新聞頻度データによる予測

朝日新聞の頻度データを用いて旧字体選好率を予測した。(5.a)式に基づいて新旧字体ペア間での頻度比の対数値を説明変数、旧字体選好率を目標変数とする回帰分析を行った結果、表2に示す(5.1)式が得られた。説明変数と目標変数の相関係数は $r=0.73$ ($p<.01$, $df=84$)、説明率は52.90%となった。説明変数と目標変数の相関図を図2に示す。

次に(5)式の予測力と比較するために、頻度を対数変換しない(6)式に基づく回帰分析を行った。

$$\text{旧字体選好率} = a \times K \times (\text{旧字体頻度} - \text{新字体頻度}) + b \quad \text{-----}(6)$$

その結果、説明変数と目標変数の相関係数は $r=0.21$ ($p<.05$, $df=84$)、説明率はわずかに4.61%であった。

表1 新旧 86 ペアの旧字体選好率 (旧字体%), 親近度差 (Fam 差), 朝日新聞頻度

ペア ID	ペア	旧字体%	Fam 差	朝日新聞頻度		ペア ID	ペア	旧字体%	Fam 差	朝日新聞頻度	
				新字体	旧字体					新字体	旧字体
1	亜亞	2.4	-1.92	1035	5	91	蠅蠅	10.6	-0.54	11	1
3	壺壺	74.1	0.96	59	20	92	竈竈	12.9	0.04	1	0
7	陥陥	8.2	-1.66	1285	0	93	条條	24.1	-0.99	9948	116
9	奥奥	1.2	-3.25	2590	0	95	嬢嬢	12.9	-1.50	108	0
11	螢螢	54.1	-0.50	98	2	96	飲飲	1.2	-3.46	2274	0
12	学學	7.1	-2.50	54725	7	100	真真	15.3	-2.16	12248	149
13	譽譽	3.5	-3.21	2198	0	101	慎慎	15.3	-2.24	2724	2
14	鶯鶯	65.9	0.87	16	4	104	楨楨	44.7	-0.37	230	2
15	鶯鶯	25.9	0.04	16	0	105	靱靱	32.9	0.09	17	2
17	会會	4.7	-2.45	161051	7	108	尽盡	2.4	-2.92	1175	2
18	桧檜	71.8	0.21	230	15	109	俛儘	30.6	-0.12	0	0
19	覚覺	1.2	-2.83	4990	3	110	数數	1.2	-3.50	29439	2
20	攪攪	10.7	-0.29	12	2	111	藪藪	40.0	-0.04	81	27
23	観觀	0.0	-3.29	7794	0	121	錢錢	9.4	-2.25	2503	1
24	灌灌	84.7	2.17	11	2	122	賤賤	65.9	1.37	2	9
29	狹狹	17.6	-2.41	948	0	125	曾曾	20.0	-1.50	926	13
31	堯堯	31.8	0.17	72	45	127	騷騷	3.5	-2.00	1619	0
32	燒燒	3.5	-3.04	2553	1	129	搜搜	5.9	-2.17	5404	0
34	区區	1.2	-3.92	28396	0	135	沢澤	38.8	-0.66	14489	643
35	欧歐	24.7	-2.59	8001	0	136	馱驛	10.6	-3.63	3315	1
38	経經	5.9	-3.00	38698	9	137	訳譯	3.5	-3.12	3161	1
39	頸頸	81.2	2.00	5	40	138	鋳鐸	45.9	0.87	0	38
44	儉儉	7.1	-2.33	36	0	139	单單	1.2	-3.83	6886	0
45	顔顔	0.0	-3.00	5078	0	140	戦戦	3.5	-3.08	37661	0
50	広廣	10.6	-2.17	19017	159	148	壳賣	2.4	-3.08	23448	0
51	拡擴	2.4	-3.66	5829	0	150	読讀	3.5	-2.45	6451	1
52	砧礪	35.3	-0.05	7	2	152	麦麥	1.2	-4.42	612	0
54	国國	10.6	-1.71	167782	170	155	発發	2.4	-2.83	57384	1
58	参參	3.5	-3.88	16992	0	164	万萬	20.0	-1.54	37713	93
59	鯨鯨	7.1	-0.96	7	0	165	砺礪	29.4	-0.83	41	0
60	贊贊	2.4	-4.17	4288	0	166	蛎蠣	30.6	-0.66	17	2
61	讚讚	4.7	-2.92	41	1	169	遙遙	35.3	-0.33	31	12
65	齒齒	1.2	-3.83	1570	0	170	謠謠	11.9	-0.91	410	0
66	齡齡	2.4	-1.99	3378	0	171	瑤瑤	23.8	0.08	1933	17
68	尔爾	37.6	1.17	2	124	172	来來	5.9	-3.96	31424	1
69	迹邇	27.1	0.16	1	11	174	竜龍	52.9	-0.63	1217	1109
74	寿壽	2.4	-2.41	1311	77	175	滝瀧	36.5	-1.12	526	75
76	涛濤	38.8	0.54	142	41	176	籠籠	81.0	1.92	22	84
77	梲檮	25.9	-0.33	5	0	177	鼠鼠	72.9	1.50	9	10
78	肃肅	7.1	-1.71	536	0	179	獵獵	4.7	-3.29	188	0
84	将將	4.7	-2.75	5408	9	180	諫諫	40.0	0.54	33	3
85	獎獎	7.1	-1.62	743	0	182	蕊藜	22.4	-0.21	0	4
90	繩繩	4.7	-2.04	2525	1	204	卷卷	14.1	-2.88	2891	0

表2 フェヒナーの法則から導出された予測モデル一覧

■ 親近度データから旧字体選好率を予測するモデル

$$\text{旧字体選好率} = a \times (\text{旧字体親近度} - \text{新字体親近度}) + b \quad \text{-----(4)}$$

$$\text{旧字体選好率} = 11.58 \times (\text{旧字体親近度} - \text{新字体親近度}) + 38.45 \quad \text{-----(4.1)} \quad r=0.86 \quad \text{説明率: 73.34\%}$$

■ 新聞頻度データから旧字体選好率を予測するモデル

$$\text{旧字体選好率} = a \times K \times \{\text{Log}(\text{旧字体頻度}) - \text{Log}(\text{新字体頻度})\} + b \quad \text{-----(5)}$$

$$\text{旧字体選好率} = 10.30 \times \{\text{Log}(\text{旧字体朝日頻度}) - \text{Log}(\text{新字体朝日頻度})\} + 41.88 \quad \text{-----(5.1)} \quad r=0.73 \quad \text{説明率: 52.90\%}$$

$$\text{旧字体選好率} = 7.40 \times \{\text{Log}(\text{旧字体読売頻度}) - \text{Log}(\text{新字体読売頻度})\} + 33.77 \quad \text{-----(5.2)} \quad r=0.70 \quad \text{説明率: 49.38\%}$$

■ 頻度差データから親近度差を予測するモデル

$$\text{旧字体親近度} - \text{新字体親近度} = K \times \{\text{Log}(\text{旧字体頻度}) - \text{Log}(\text{新字体頻度})\} \quad \text{-----(7)}$$

$$\text{旧字体親近度} - \text{新字体親近度} = 0.91 \times \{\text{Log}(\text{旧字体朝日頻度}) - \text{Log}(\text{新字体朝日頻度})\} + 0.33 \quad \text{-----(7.1)} \quad r=0.87 \quad \text{説明率: 74.82\%}$$

$$\text{旧字体親近度} - \text{新字体親近度} = 0.66 \times \{\text{Log}(\text{旧字体読売頻度}) - \text{Log}(\text{新字体読売頻度})\} - 0.36 \quad \text{-----(7.2)} \quad r=0.85 \quad \text{説明率: 72.60\%}$$

注1) 傾き a ならびに a×K, 切片 b の係数に関する検定結果はすべて有意であった (p<.01)。

注2) 対数計算においては頻度データに 1 を加算した。

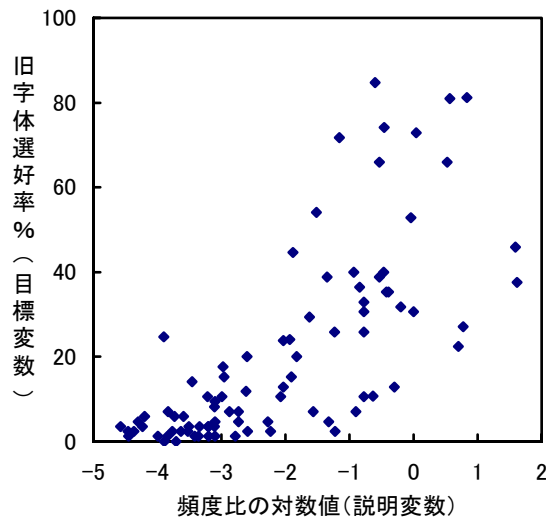


図2 新旧ペア間における頻度比対数値（説明変数）と
旧字体選好率（目標変数）の相関図【 $r=0.73$ 】

同様にして、読売新聞の頻度データから旧字体選好率を予測した。(5.a) 式の回帰分析から (5.2) 式が得られた。説明変数と目標変数の相関係数は $r=0.70$ ($p<.01$, $df=84$)、説明率は 49.38%であった。次に (6) 式による予測を行ったところ、説明変数と目標変数の相関係数は $r=0.26$ ($p<.01$, $df=84$)、説明率は 6.72%となった。ここでも (6) 式のモデルは予測力が著しく劣ることが確認されたため、以後の分析では (6) 式に基づく回帰分析は扱わないことにした。

3. 3. 頻度差データから親近度差を予測するモデル

先に述べた (3) 式から (3.a) 式を減じると (7) 式が導かれる。

$$\text{旧字体親近度} = K \times \text{Log}(\text{旧字体頻度}) + C \quad \text{-----}(3)$$

$$\text{新字体親近度} = K \times \text{Log}(\text{新字体頻度}) + C \quad \text{-----}(3.a)$$

$$\text{旧字体親近度} - \text{新字体親近度} = K \times \{ \text{Log}(\text{旧字体頻度}) - \text{Log}(\text{新字体頻度}) \} \quad \text{-----}(7)$$

目標変数の値を漢字 1 文字ごとの親近度から、説明変数の値を朝日新聞頻度データからそれぞれ計算し、回帰分析を行った。その結果、表 2 に示す (7.1) 式を得た。説明変数と目標変数の相関係数は $r=0.87$ ($p<.01$, $df=84$)、説明率は 74.82%であった。同様にして

読売新聞頻度データから (7.2) 式が得られた。説明変数と目標変数の相関係数は $r=0.85$ ($p<.01$, $df=84$)、説明率は 72.60% であった。

3. 4. 予測モデルの相互関係：傾きと切片の比較対照

旧字体選好率に関する (4.1) 式の a 項は 11.58、朝日頻度データによる (7.1) 式の K 項は 0.91 であるから、朝日頻度データによる (5.1) 式の $a \times K$ 項は $11.58 \times 0.91 = 10.54$ だと予測できる。この値と実際の (5.1) 式の $a \times K$ 値 10.30 と比較すると、誤差は 0.24 と非常に小さいものであった。読売頻度データについても、旧字体選好率に関する (4.1) 式の a 項は 11.58、(7.2) 式の K 項は 0.66 であるから、読売頻度データによる (5.2) 式の $a \times K$ 項は $11.58 \times 0.66 = 7.64$ だと予測できる。この値と実際の (5.2) 式の $a \times K$ 項の値 7.40 と比較すると、誤差は 0.24 であった。

このように朝日新聞と読売新聞のいずれにおいても説明変数と目標変数の相関は極めて高かった。これらの結果から、本研究のモデルは相互に整合性を有し、基準関連妥当性を満たすと言えよう。

4. 考 察

新聞漢字頻度から字体選好課題の結果を予測するためにフェヒナーの法則から予測モデルを導出した。それらのモデルは、朝日新聞データの場合、説明率が 50% 以上で説明変数と目標変数の相関係数が 0.73 を超え、かなり良好な予測精度を示すことが明らかになった。きわめて単純なモデルでありながら、しかも新聞漢字頻度の要因だけでこの説明率に到達するという事実は興味深い。

この結果は、先に述べた単純接触効果でうまく説明できる。新聞漢字頻度データが言語生活全般における漢字使用の傾向を適切に代表しているのであれば、その指標から人間が個々の漢字に接触する頻度を推測できるはずである。社会でよく使われる漢字は高い確率で目に入り、あまり使われない漢字は目に入る確率が低いと考えられる。それゆえ、ある漢字を目にする確率（接触頻度）は、その漢字の使用頻度と密接な関係にある。新奇刺激を繰り返し経験するだけでその刺激に対する選好度が高まる (Kunst-Wilson & Zajonc, 1980) ことから、接触頻度の高い漢字はそうでない漢字よりも選好度が高くなるのであろう。(当然のことながら、人間にとって漢字も当初は新奇刺激である。)

Zajonc (1968), Kunst-Wilson & Zajonc (1980) をはじめとする多くの先行研究は、新奇刺激を繰り返し経験すると「なじみ」が生じて親近度が高くなり、その影響で選好度が上

昇すると説明している。本研究でもその説を裏付ける結果が得られた。字体選好率に対する予測力がもっとも優れていたのは新旧字体ペア間の親近度差を説明変数とするモデルであったが、これは字体選好の判断に際して、被験者が新旧字体間で親近度を比較し、より親近度の高い字体を選択するという心的プロセスを反映した結果だと考えられる。

字体選好課題の結果は新聞漢字頻度だけで 100%説明できるわけではない。残差が 50%ほどあるのだが、この残差が生じた原因として次の 3つの可能性を指摘できる。

第 1 に、言語生活での漢字接触頻度を代表する指標として新聞漢字頻度データが妥当なのかという疑問が残る。笹原・横山・ロング (2003) が指摘するように、漢字は新聞だけではなく、書籍、ちらし、テレビのテロップや映画の字幕、携帯メールやパソコン、書道での字、個人の手書き、看板、街で見かける字など、学校で習う文字や辞書に載っている字以外にも、さまざまなものが存在することを忘れてはいけない。「人間はどのような漢字をどのくらい目にしているのか」あるいは「人間はどのような漢字を好むのか」という問いに正確に答えるには、我々を取り囲む「漢字環境」を科学的に観測・調査し、その実態を十分に知り尽くす必要がある。新聞、書籍、看板、携帯メールなど各種のメディアに登場する漢字の使用状況を綿密に調査し、社会における漢字の使用頻度を明らかにできた段階で漢字接触頻度をかなり正確に予測できるようになるだろう。字体の選好にはいろいろな原因が考えられる。そのうち新聞がどのくらいの比重を占めているかを正確に知るには、多様な文字媒体の実態を反映したコーパスが必要である。しかし、現時点ではそのようなコーパスは存在しないため、今後の研究に期待したい。

第 2 に、字体選好判断においては出現頻度以外の要因も大きな役割を果たすことがあるのではないかという点がある。例えば、読み手に配慮した表記の選択を行う場面を想像してみると、読み手の年齢や教育歴などが書き手側にあらかじめ分かっている場合は読み手の社会的属性に基づいた推論がなされて適切な表記が選択されると考えられる。この種の語用論的な要因は本研究のモデルには含まれていない。残差 50%のうち、語用論的要因で予測可能な部分がどの程度あるかについてはこれからの研究課題である。

第 3 に、字体選好課題の結果を予測するのに有効な説明変数が、使用頻度の他にもあるのではないかという点である。漢字刺激は、画数、対称性、複雑性など、多くの属性を有する。これらの変数を取り込んだ多変量解析（重回帰分析など）が必要であろう。この点についてもこれからの課題である。

本研究で提唱された字体選好予測モデルは、文字生活のある側面とフェヒナー法則の関

係について定式化したものとみなせる。また、この予測モデルは、日常の漢字認知に潜在する単純接触効果の大きさを推定する手法だとも言える。今回の選好判断実験に参加した女子大学生 85 名全員が朝日新聞（あるいは読売新聞）を日常的に読んでいるわけではないことから、選好と新聞における異体字頻度の間に直接的な因果関係を想定するのは不自然である。しかしながら、選好と新聞データとの間にかなり高い相関が見られるということは、新聞はさまざまな文字媒体における文字使用の傾向をある程度は代表している媒体だと考えても大過ないと解釈できよう。

今後は、新聞だけではなく雑誌などの漢字調査データを使うとともに、先に指摘したように重回帰分析による説明変数の取舍選択が必要となるであろう。さらに、「ひんしゆくー顰蹙」などの語レベルの表記選好についても、本研究の予測手法がどのくらい有効なのかを検討すべきである。表記別の頻度がコーパス等から求まりさえすれば、人間がどちらの表記を選ぶかを数量的に予測できる。その精度を検証する研究が今後登場することを期待したい。

謝辞

本研究で使用した字体選好課題のデータは、笹原宏之氏（早稲田大学社会科学部）と横山が共著で公刊した先行論文ならびに CD-ROM に基づくものである。データ使用の許諾をくださった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）

- 朝日新聞社（1994）『CD-HIASK'93 朝日新聞記事データベース』、紀伊國屋書店・日外アソシエーツ
- 文化庁文化庁国語課（2000）『漢字出現頻度数調査(2)』漢字字体関係参考資料集、文化庁
- 文化庁文化庁国語課（2005）『平成 16 年度 国語に関する世論調査』、文化庁
- Chikamatsu, N., Yokoyama, S., Nozaki, H., Long, E., & Fukuda S. (2000) A Japanese Logographic Character Frequency List for Cognitive Science Research. *Behavior Research Methods, Instruments, and Computers*, No. 32, Vol. 3, pp.482-500, Psychonomic Society.
- Elliot, R., & Dolan, R. (1998) Neural Response during Preference and Memory Judgments for Subliminally Presented Stimuli: A Functional Neuroimaging Study. *The Journal of Neuroscience*,

- 1998, 18, pp.4697-4704, Society for Neuroscience
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science*, 207, pp.557-558.
- Moin, B. (2003) The Warm Glow Heuristic: When Liking Leads to Familiarity. *Journal of Personality and Social Psychology*, 85, pp.1035-1048
- NTT コミュニケーション科学基礎研究所〔監修〕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, 三省堂
- 笹原宏之・横山詔一(1997)「大学生による異体字の選択行動」『計量国語学』21 巻 3 号 pp.117-118 (計量国語学会第 41 回大会発表要旨)
- 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
- 横山詔一・笹原宏之・野崎浩成・エリク＝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

(2005 年 11 月 29 日受付)

(英文アブストラクト)

PAPER

Can We Predict Preference for Kanji Form from Newspaper Data on Character Frequency?

YOKOYAMA Shoichi (National Institute for Japanese Language)

Descriptors: Kanji variant, preference, familiarity, Fechner's law, mere exposure effect, Two-Alternatives Forced Choice

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 psychophysical model, based on Fechner's law, which predicts familiarity judgments performance from frequency of kanji variant. The introduced model was tested on empirical data obtained from an experiment, in which 85 subjects were presented 263 traditional/simplified variant pairs and asked which form they would prefer to use in word processing. Regression analyses were conducted with the preference ratio for the traditional form as the dependent variable, and the difference of the log-frequency of the characters as the independent variable (cf. Formula 5.1) and results showed significant correlations ($r = .73$ for Asahi and $r = .70$ for Yomiuri). The data suggest the effect of frequency, supporting the so-called "mere exposure effect" theory by Zajonc (1968) in Japanese kanji processing.

(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)

日本語の漢字には「異体字」の豊富なバリエーションが存在する。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文字の集合を指す。本研究は、新聞漢字頻度データから「字体選好課題」の結果がどのくらい正確に予測できるかを検討した。字体選好課題とは、ワープロやパソコンを打っている場面で異体字ペアのどちらをより使いたいか被験者に選択させる課題である。263組の異体字ペアを85名の被験者に呈示したデータを分析に用いた。精神物理学の分野で有名なフェヒナーの法則 (Fechner's law) を出発点として、旧字体頻度と新字体頻度の比を求め、それを対数変換した説明変数から旧字体選好率を予測する回帰式を求めた。その結果、目的変数と説明変数の相関は $r=.73$ に達することが示された。これを接触相対性モデル (exposure relativity model) と名付けた。