

字体選好の地域差に関する計量的研究

横山 詔一

1. はじめに

1.1. 異体字問題の影響

現代社会の文字コミュニケーションに関する手がかりを得るために、「異体字 (kanji variant)」の選択行動に着目した統計的分析を行った。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文字の集合を指す。異体字のバリエーション (変異) の豊富さは現代日本の文字生活になんらかの影響をもたらしているものと推測される。しかし、その詳細な実態はほとんど解明されていない (笹原・横山・ロング、2003)。そこで、ワープロや携帯メールなどの情報通信機器 (IT 機器) を使って漢字を書いているとき、変換候補として「桧-檜」などの異体字ペアが示されるとどちらの字体を人間は選択するのか、という実験心理学的な方法を用いてデータを収集した。

日本語の異体字をめぐる IT 機器環境は、2007 年ごろから変化しつつある。たとえば、2007 年 1 月までは、一般的な文書作成ソフトで図 1 (b) の「葛 (L+人)」を使いたい場合であっても、原則として図 1 (a) の「葛 (ヒ)」しか使えなかった。Web での表示でも同様であった。

葛 葛

(a) (b)

図-1 JIS 漢字規格で例示字形が変更された例

ところが、2007 年 1 月に発売された Windows OS 「ビスタ (VISTA)」で日本語フォントの約 150 字について字体が変更され、その影響により多くの IT 機器で図 1 (b) 「葛 (L+人)」が使えるようになった。その反面、かつては使用できた図 1 (a) 「葛 (ヒ)」は IT 機器で使えない場合が増えている。

2012 年現在では、2007 年以前に作成したテキストファイルを IT 機器で表示・印字すると、図 1 (a) の「葛 (ヒ)」が図 1 (b) の「葛 (L+人)」になって表示・印字されることが少なくなっている。本稿で「葛 (ヒ)」や「葛 (L+人)」という注記を付しているのは、この字体入れ替わり問題に対する予防策である。

マイクロソフト社がピスタで日本語フォント約 150 字の字体を変更した理由は、JIS 漢字規格の一つである「JIS X 0213:2004」に対応するためであった。周知の通り、JIS 漢字規格は漢字政策の一つとして強い影響力を持つ。この JIS X 0213:2004 とは、経済産業省が JIS 漢字 (JIS X0213) の 160 字あまりについて印刷標準字体 (いわゆる康熙字典体) にしたがって規格書の例示字形を 2004 年 2 月に変更したものを指す。IT 関連企業は、そのうちの約 150 字の異体字に対応したとされる。

本研究は、この異体字問題を念頭において 2007 年 1 月以前に観測された文字意識の地域差を計量的に解析し、文字環境モデル (図 3: 後述) の観点から考察を試みる。以下、図 1 のように異体字ペアの字形を画像で示す場合は、日本規格協会が国立国語研究所および情報処理学会と共同でデザインした約 6 万字の平成明朝体グリフ (文字フォントの字母のようなもの) から笹原・横山・ロング (2003) による規準にしたがって選んだ。その例を図 2 に示す。また、調査の対象者は、調査実施時期の 2005 年 12 月当時に、すでにインターネットを日常的に使用していた層からランダム抽出された人であった。

| | | | | | |
|----|---|---|----|---|---|
| 3 | 壺 | 壺 | 41 | 倦 | 倦 |
| 6 | 秤 | 秤 | 42 | 捲 | 捲 |
| 11 | 蚩 | 蚩 | 43 | 齷 | 齷 |
| 14 | 鶯 | 鶯 | 44 | 儉 | 儉 |
| 17 | 会 | 會 | 45 | 顔 | 顔 |
| 18 | 檜 | 桧 | 46 | 諺 | 諺 |
| 21 | 葛 | 葛 | 54 | 国 | 國 |
| 22 | 喝 | 喝 | 55 | 摑 | 摑 |
| 23 | 観 | 觀 | 57 | 薩 | 薩 |
| 24 | 灌 | 灌 | 63 | 祇 | 祇 |
| 28 | 俠 | 俠 | 64 | 桷 | 桷 |
| 29 | 狭 | 狭 | 68 | 爾 | 尔 |
| 30 | 頬 | 頰 | 69 | 迓 | 邇 |
| 31 | 堯 | 堯 | 72 | 溺 | 溺 |

図-2 異体字ペアを平成明朝体グリフで印字した例

1.2. 字体選好と単純接触効果の関係

図1の文字が日常生活に直接関係する地域としては、東京都葛飾区、東京都江戸川区葛西・西葛西周辺、奈良県葛城市（市名の正式表記は「葛（ヒ）」）、岩手県葛巻町などが考えられる。それらの地域では、住民の多くが長期間にわたって繰り返し「葛（ヒ）」もしくは「葛（L+人）」の字体・字形に意識的・無意識的に接触していると想像される。その結果、他の地域とは違った文字意識が芽生えている可能性がある。その違いとは、どのようなものなのだろうか。

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、それが心内辞書（mental lexicon）の形成や言語行動に影響を与える（横山、2006a）。図3は文字環境研究の観点から文字生活を捉えた循環モデルの図である。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。

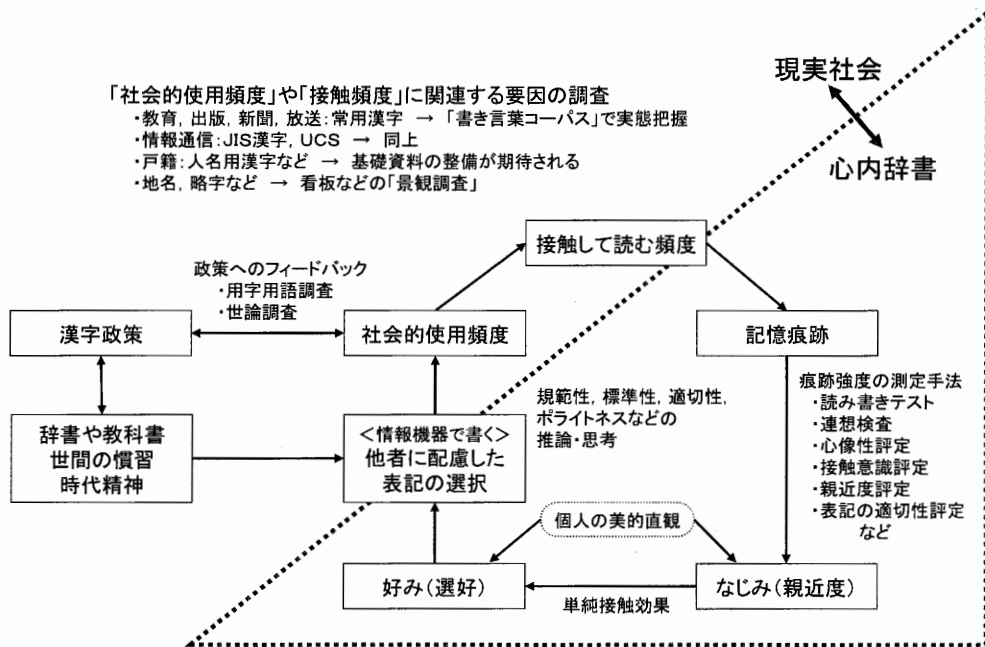


図-3 文字環境の循環モデル：横山（2006a）を拡張・改訂

漢字心理は人間が漢字を読む（識別や包摂も含む）場合だけではなく、漢字を使用する際にも重要な要因となり、それが社会的な使用頻度に波及する。日本語教育学における漢字習得の問題なども、大局的に見れば図3の「漢字政策」や「社会的な使用頻度」の部分に位置づく。

地名の漢字が地域住民の文字意識に及ぼす影響について、横山（2006b）、Yokoyama & Wada（2006）は「単純接触効果（mere exposure effect）」が重要な役割を果たしているとの仮説を提唱した。単純接触効果とは社会心理学の用語で、「なじみ（親近度：familiarity）」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「好意度 (favorability)」が高まる現象を指す (Zajonc, 1968 ; Kunst-Wilson & Zajonc, 1980 ; Elliot & Dolan, 1998 ; 横山・渡邊, 2007)。

この仮説を検証するため、横山 (2006b) は、「葛 (ヒ) - 葛 (L + 人)」ペアを東京都葛飾区住民 (葛飾群) 60 名と山の手地区住民 (山の手群) 57 名に呈示して、より使いたい方の字体を選択するよう求めた (2005 年 12 月に実施)。より使いたいと判断された字体は、より好まれている字体だと解釈して大過ない。以下、このような課題を「字体選好課題」という。

2005 年 12 月時点での字体選好課題の結果を表 1 に示す (横山, 2006b)。文字論の通説に従って「葛 (ヒ)」を新字体、「葛 (L + 人)」を旧字体とした。葛飾群は旧字体を選んだ人数の割合が新字体の約 1.9 倍に達した。逆に、山の手群は新字体を選んだ人数の割合が旧字体の約 1.7 倍になった。葛飾群と山の手群では人気のある字体に違いがあることが示された。カイ 2 乗検定の結果、1%水準で有意差が見られ ($df = 1$)、地域によって字体選択パターンが逆転していることが分かった。「葛」ペアについては、葛飾群は山の手群よりも旧字体を好むことが明らかになった。

表-1 「葛」字体選好の地域差 (数値は人数、カッコ内は%)

| | 新字体 | 旧字体 | N |
|------|-----------|-----------|-----------|
| 葛飾群 | 21 (35.0) | 39 (65.0) | 60 (100) |
| 山の手群 | 36 (63.2) | 21 (36.8) | 57 (100) |
| N | 57 (48.7) | 60 (51.3) | 117 (100) |

2. 字体選好データに基づく居住地の推測

2.1. 居住地と字体選好の因果関係

ここまでは【原因：居住地での文字接触】→【結果：地名の字体選好】という関係を検討してきた。表 1 のカイ 2 乗検定は、その目的に合致している。しかし、カイ 2 乗検定だけではデータから有用な情報を十分に取り出すことが難しい。多変量解析を利用すれば、結果から原因を推測する、つまり【結果：地名の字体選好】→【原因：居住地での文字接触】という新たな視点に立つ分析も可能になる。

方言研究に眼を転じると、話者の音声データから方言研究者が話者の居住地を的中させたという例は枚挙にいとまがない。これは【結果：話者の音声選好】→【原因：居住地での音声接触】という推論の図式に近いのではないか。同じく、文字表記に関するデータから回答者の居住地を的中させることが可能であるならば、言語の空間変異をめぐる新たな手がかり・知見を得る契機になるかもしれない。

いくつかの異体字ペアについて字体の好みに関する質問を行い、その回答から葛飾住民か山の手住民かを当て推量 (勘) よりも高い精度で判別できるならば、文字意識にも地域差が存在する計量的な証左になる。当て推量だと、居住地を予測した場合の的中率は 50%前後にとどまる。

2.2. ロジスティック回帰分析の導入

字体選好データから居住地を推測する手法として多変量解析の一つである「ロジスティック回帰分析 (logistic regression analysis)」を導入する。この分析では、字体選好データを説明変数に、居住地 (カテゴリー変数) を目的変数とし、判別分析と同様の多変量解析をおこなう。

ロジスティック回帰分析は、医学分野では、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になる危険性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かを検討する研究などでよく使われる。情報通信技術の分野では、迷惑メール (いわゆるスパムメール) か否かを自動的に判別するメールソフトの開発などに利用されている。目的変数が「生存/死亡」のように2値をとるデータに適用するケースが多い。

ロジスティック回帰分析は、目的変数によって以下の2通りの使い方がある。

- (1) 目的変数がカテゴリー (名義尺度) の場合。
- (2) 目的変数が比率データの場合。確率なら0から1、百分率 (%) なら0から100の範囲に数値が収まる。

本研究は前者の分析をおこなった。目的変数は「葛飾区/山の手地区」の2値変数で、説明変数は後述の会場実験で採集した字体選好データであった。108組の異体字ペアについて選好データを収集し、それらを説明変数として分析に投入した。選好データは「新字体/旧字体」のいずれを選択したかの2値変数。そのうち、新旧いずれかの字体だけによる例を表2に示す (2012年時点のIT機器で表示できない字体が少なくないため)。

表-2 説明変数の例 (異体字ペアのうち新旧いずれかの字体を掲出)

壺、秤、螢、鶯、桧、葛、喝、観、漉、侠、狭、頰、堯、欧、軀、鷗、頸、倦、捲、鹵、諺、搨、薩、祇、榊、尔、迨、溺、颯、寿、禱、涛、梲、哨、屑、囊、鉛、餅、真、慎、填、顛、楨、尽、俛、藪、鯖、鑄、滯、卿、巽、撰、錢、賤、煎、曾、噌、樽、啄、塚、沢、駟、訳、釵、蟬、驛、箆、洩、澆、醜、稗、瀨、万、砺、蛭、遥、謠、瑤、竜、滝、箆、鼠、蠟、獵、諫、剥、蔑、隄、襄、蜻、琢、撥、喻

このようなデータに対して、これまでは「判別分析」の適用を第一に考えることが多かったと思われるが、判別分析は説明変数が多変量正規分布していることを前提としているため、本研究のデータに使うのは好ましくない。また、「林の数量化理論Ⅱ類」による分析も可能ではあるが、説明変数の統計的有意性が検定できないため、最近の医学や経済学ではほとんど使用されていない。

ちなみに、ロジスティック回帰分析を初めて言語研究に導入したのは、Labov (1972) だと言われている。日本語に関する研究では、Hibiya (1988) と Matsuda (1993) の研究が先駆的である。最近では、音声共通語化の研究で横山・真田 (2010) がある。文字生活研究での利用は、横山 (2006a) と Yokoyama & Wada (2006) が嚆矢であろうか。

2.3. データ収集の方法

字体選好データの収集方法は次の通りであった。先に示した表1は、この調査で得た資料のうち「葛(ヒ) - 葛(L+人)」ペアだけを速報的に分析した結果である。

地域差の設定

葛飾群は、親の代から葛飾に住んでいて本人が葛飾生まれ葛飾育ちの生え抜き住民のほか、居住歴が10年以上もしくは5年以上の層もターゲットに含めた。一方、山の手群は葛飾区や江戸川区に親戚や友人がおらず、その周辺を訪れたこともない層を抽出した。

年代差の設定

年齢要因は20代、30代、40代、50代とした。葛飾群と山の手群で年齢条件がなるべくそろうようにした。調査参加者は全員女性であった。

調査手続き

「パソコンなどで字を打っているときにより使いたいと感じる方の字を選んでください」と教示し、IT機器で字を書く場面をイメージしたとき、異体字ペアのうち、より使いたいと感じる方の字体を2肢強制選択法(2-Alternatives Forced Choice)で直観的に選ばせた。この手法は、JIS漢字に含まれない字体(字形)であってもパソコンで表示・印字できるよう文字グリフ集合(後述の電子政府用文字グリフ集合)を準備して紙ベースで調査や実験を行えることから、IT機器における異体字の利用制限がない。

文字刺激の統制

調査に用いた異体字ペアは、JIS X 0213:2004による字形変更などを考慮して横山・笹原・當山(2006)が用いた263ペアから108ペアを抽出したものであった。新旧字体の左右位置は調査参加者ごとにランダム化した。

文字のデザインは日本規格協会による平成明朝体を用いた。調査用紙に印字した文字刺激は、国立国語研究所・日本規格協会・情報処理学会が共同プロジェクトで制作した「電子政府用文字グリフ集合(約6万字)」から抽出したものであった。これらは国の行政情報処理で使用する漢字デザインの標準であり、漢字字形の原器といえる。新旧ペアは笹原・横山・ロング(2003)による次の規準にしたがって選ばれた。

- ・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。

上記の規準に適合した異体字集合から、さらに、調査参加者になじみがないと思われる字や、字種が多いグループの字を原則として削除した。また、パソコンに搭載されていないフォントを扱う必要があるため、電子政府用文字グリフをPDFファイル化することで、パソコンを用いて紙に印字できるようにし、調査票の冊子を作成した。フェイスシートには氏名、年齢、ワープロ使用経験の有無などについての回答欄も設けた。

場所と日時

2005年12月に葛飾区金町と中野区中野坂上に調査会場を設け、調査参加者に集まってもらった。

2.4. 結果と考察

以下のロジスティック回帰分析に用いたデータは、葛飾群 58 名と山の手群 57 名の合計 115 名であった（葛飾群 60 名のうち、異体字ペアの一部分の判断に欠損データを含む 2 名を分析から除外した）。変数にかかる係数の推定は Yokoyama & Wada (2006) と同じく最尤推定法によっておこなった。計算は SPSS を用いた。

ロジスティック回帰分析で変数増加法によるステップワイズ法を行い、ワルド (Wald) 検定により 5%水準で統計的に有意な説明変数を抽出したところ、有意な説明変数は全部で 5 つあった。それらの係数を表 3 に示す。

表-3 ロジスティック回帰分析で得た係数と的中率 (%)

| 説明変数 | 1 変数 | 2 変数 | 3 変数 | 4 変数 | 5 変数 |
|-------------------|---------|---------|---------|---------|---------|
| 葛 (ヒ) - 葛 (L + 人) | 1.181 | 1.406 | 1.358 | 1.354 | 1.432 |
| 堯 - 堯 | | - 1.327 | - 1.855 | - 1.772 | - 2.384 |
| 頸 - 頸 | | | 1.337 | 1.318 | 1.082 |
| 藪 - 藪 | | | | - 1.549 | 1.224 |
| 塚 - 塚 | | | | | - 1.686 |
| 定数 | - 0.588 | - 0.418 | - 0.570 | - 0.441 | - 0.821 |
| 葛飾の的中率 | 65.5 | 55.2 | 69.0 | 67.2 | 79.3 |
| 山の手の中率 | 63.2 | 80.7 | 71.9 | 77.2 | 68.4 |
| 全体的中率 | 64.3 | 67.8 | 70.4 | 72.2 | 73.9 |

注)「塚」の異体字は後述の表 7 の下「23」で字形を示す。

例えば、3 変数の係数で居住地を推定するやり方を 3 変数モデルという。推定の手順は次の通りである。まず、式【1】を計算する。

$$Z = 1.358 \times \text{葛} - 1.855 \times \text{堯} + 1.337 \times \text{頸} - 0.570 \quad \text{【1】}$$

ある調査参加者が「葛、堯、頸」のそれぞれについて、新字体を選択した場合は 0 を、旧字体を選択した場合は 1 を代入して Z の値を得る。その Z を式【2】のロジスティック関数に代入すると、その人が葛飾群である推定確率 p_1 が計算できる。

$$p_1 = 1 / [1 + \exp(-Z)] \quad \text{【2】}$$

推定確率 p_1 が 0.5 (つまり 2 分の 1) 以上であれば葛飾群、0.5 以下であれば山の手群と判別する。葛飾群と山の手群は排反事象なので、山の手群である確率は $1 - p_1$ になる。つまり、 p_1 はその人の「葛飾群らしさ」を確率で表現した指標になっている。

以上による 3 変数モデルの的中率は 70.4% に達した。表 3 のように葛飾群と山の手群での的中率がほぼ等しく、バランスがとれている。2 変数、4 変数、5 変数モデルは、どれも葛飾群と山の

手群での中率に相当な差があり、アンバランスであった。また、1変数モデルは全体的中率が65%を下回った。これらを総合して判断すると、3変数モデルが適当だと考えられる。

ただし、3変数モデルを採用するとして、「堯-堯、頤-頤」の2ペアが果たす役割については解釈不能である。解釈に資するデータがまったく存在しない。その点で、1変数モデルでも的中率が64%以上に達するのであるから十分だという考え方も成立するだろう。的中率を少しでも向上させたいのか、それとも現象の説明や解釈に重心を置くかで、1変数モデルと3変数モデルのいずれかを選択することになるだろう。

このあたりの事情は、最近、自然言語処理分野できわめて発展が顕著な「統計的機械学習モデル」と似た部分がある。統計的機械学習モデルは係数の解釈が人間にはできないようになっている。

次に、データを収集した異体字108ペアに対して異体字選好に地域差が見られるかカイ2乗検定を行った。その結果、「葛」のほかには「堯-堯」と「囊-■」の2ペアにおいて5%水準で有意差が認められた(いずれも $df = 1$)。「囊-■」ペアを図4に示す。

囊 囊

(a) (b)

図-4 「囊」の異体字ペア

「堯-堯」ペアの結果を表4に示す。文字論の定説にしたがってペアの「堯」を新字体、「堯」を旧字体という。葛飾群は山の手群よりも新字体を好む傾向がある。これは「葛」ペアで葛飾群が旧字体を好んだ傾向とは逆になっている。この理由を究明するにはデータや資料が不足している。カイ2乗検定を個々の異体字ペアに適用したため、全体で108回の繰り返し分析になった。その関係で生じた検定の過誤かもしれない。

表-4 「堯」字体選好の地域差 (数値は人数、カッコ内は%)

| | 新字体 | 旧字体 | N |
|------|-----------|-----------|-----------|
| 葛飾群 | 52 (86.7) | 8 (13.3) | 60 (100) |
| 山の手群 | 40 (70.2) | 17 (29.8) | 57 (100) |
| N | 92 (78.6) | 25 (21.4) | 117 (100) |

「囊-■」ペアの結果を表5に示す。ペアの「囊」を新字体、もう一方を旧字体という。「堯」ペアと同じく、葛飾群は新字体を好む傾向があり、「葛」の場合とは逆のパターンであった。この理由は今のところ不明であるが、少なくとも葛飾群は山の手群よりも旧字体を好む傾向が強いわけではないようである。むしろ、「葛」を例外として、山の手群の方が葛飾群よりも旧字体を好む傾向にあると言えよう。

表-5 「藁」字体選好の地域差（数値は人数、カッコ内は%）

| | 新字体 | 旧字体 | N |
|------|-----------|-----------|-----------|
| 葛飾群 | 42 (70.0) | 18 (30.0) | 60 (100) |
| 山の手群 | 28 (49.1) | 29 (50.9) | 57 (100) |
| N | 70 (59.8) | 47 (40.2) | 117 (100) |

先にも述べたように、JIS 漢字規格の関係で、2007年1月ごろまでのパソコンなどのIT機器では「葛（ヒ）」は表示できても「葛（L+人）」は表示できないケースが少なくなかった。国民各層がIT機器を使う時代に入ったことから、全国規模で見るとどちらかと言えば「葛（ヒ）」になじみを感じるという人が2007年1月ごろまでは主流になっていた。一方、東京都葛飾区や江戸川区葛西・西葛西周辺は、他の地域よりも相対的に「葛（L+人）」を目にする機会が2007年1月以前から多いようである。葛飾区の路上で看板や住所表示板を観察してみると、確かに「葛（L+人）」がすぐに目に飛び込んでくる。また、葛飾区役所の公式ホームページでも区名の正式な表記は「葛（L+人）」だと注記されていた（2007年4月時点）。

以上の結果から、次の解釈が考えられる。2007年1月以前から葛飾区で日常生活を過ごした人は、自然に「葛（L+人）」に対する単純接触効果が醸成され、2007年1月以前においても「葛（ヒ）」よりも「葛（L+人）」を好む人の割合が他の地域に比べて高くなっていてはいないか。同じ東京都内でも葛飾地区住民（葛飾群）と山の手地区住民（山の手群）を2007年1月以前のデータで比べてみると、「葛」字体の好みの傾向に違いがあり、葛飾群の方が「葛（L+人）」を好む人が多かったという結果は、この予想を裏付けるものだと考えられる。

なお、「葛（ヒ）」あるいは「葛（L+人）」を日常生活で目にするのは、葛飾区などの地名、葛西選手などの人名、葛西駅などの駅名、葛を用いた食品名や屋号、葛根湯などの生薬成分名が大部分を占めており、それらを除くと「葛藤」などの語があるぐらいで、全国を見渡すと一般の人が接触する頻度が高い文字の部類には入らない。よって、山の手群の結果は、2007年1月以前の全国の一般的な傾向を反映したものと考えられる。

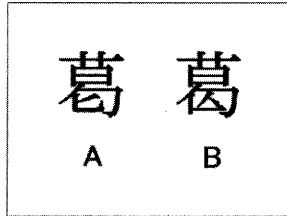
3. ネット調査による追試

ここでは表1の結果が他の調査手法でも再現できるかを確認するため、経営学の分野などで市場調査によく使われるネット調査を用いて追試した（横山・高田・當山・米田、2009）。

3.1. データ収集の方法

単漢字の異体字ペアのほか「森鷗外」などの固有名詞や「剥製」などの一般名詞のペアも取り上げ、25ペアについてインターネットを介してWeb画面に呈示した。画面イメージの例を図5に示す。

【Q.1-S.25】AとB, どちらの字を使いたいが、教えてください。(ひとつだけ)【必須】



- C 1. Aを使う
- C 2. Bを使う

図-5 Web による調査画面の例

地域差の設定

東京都葛飾区ならびに江戸川区の住民（葛飾・江戸川群）と、東京都山の手ならびに神奈川県（山の手・神奈川群）を調査対象とした。江戸川区を加えたのは、葛西や西葛西などの地名があるためであった。ランダムサンプリングは野村総合研究所が構築した全国約 20 万人の調査協力者データベースを用いておこなった。

年代差の設定

20 代、30 代、40 代がそれぞれ 100 名ずつ、50 代は 89 名で計 389 名。全員女性であった。50 代以外は葛飾・江戸川群と山の手・神奈川群で 50 名ずつを割り当てた。50 代は葛飾・江戸川群が 39 名、山の手・神奈川群は 50 名であった。

調査手続き

異体字ペアを Web 画面に呈示して、「パソコンなどで字を打っているときにより使いたいと感じる方の字」を調査参加者に強制選択させた。

調査実施時期

2006 年 3 月であった。

文字刺激の統制

Web 画面に呈示した文字刺激は、国立国語研究所・日本規格協会・情報処理学会が共同プロジェクトで制作した電子政府用の文字グリフ集合（約 6 万字）から抽出したものであった。

3.2. 「葛」ペアの結果

「葛」の新旧字体について、全データを対象にした選好課題の結果を表 6 に示す。葛飾・江戸川群は山の手・神奈川群に比べて旧字体を好む割合が高く、全体的には横山（2006b）と同様の傾向にあることが確認された。ただし、表 1 と比較すると、表 6 は新字体の選好率が 10% 程度高くなっている。表 6 の数値については、ネット調査によるサンプルのバイアス効果のほか、葛飾・江戸川地区の居住歴が短い層を排除しなかった影響もあったと考えられる。それに対して表 1 のサンプルは、葛飾地区の生え抜きを中心に居住歴が 10 年以上あるいは 5 年以上の層を調査参加者とした。

カイ 2 乗検定の結果、1% 水準で有意差が見られ、葛飾・江戸川群の方が山の手・神奈川群よ

りも「葛 (L + 人)」を好むことが示された ($df = 1$)。これは先に述べた表 1 の結果と一致する。

表-6 「葛」字体選好の地域差追試 (数値は人数、カッコ内は%)

| | 新字体 | 旧字体 | N |
|----------|------------|------------|-----------|
| 葛飾・江戸川群 | 89 (47.1) | 100 (52.9) | 189 (100) |
| 山の手・神奈川群 | 145 (72.5) | 55 (27.5) | 200 (100) |
| N | 234 (60.2) | 155 (39.8) | 389 (100) |

調査実施時期 (2006年3月) の葛飾区役所の公式ホームページでは、区名の正式な表記は図 5 の B の「葛 (L + 人)」であることが注記されていた。その当時から葛飾・江戸川地域の看板や街頭の地名表示に図 5 の B の「葛 (L + 人)」が多く存在した可能性がある。

3.3. その他のペアの結果

表 7 にネット調査で用いた項目全体 (25 ペア) を示す。個々のペアについて、地域差がみられるかカイ 2 乗検定をおこなった。その結果、上で述べた「15. 葛 (ヒ) - 葛 (L + 人)」を除いて、いずれも有意ではなかった。

表-7 ネット調査の項目リスト

1. 観-観、2. 灌-灌、3. 会-會、4. 桧-檜、5. 経-經、6. 頸-頸、7. 亜-亞、
8. 壺-壺、9. 竜-龍、10. 籠-籠、11. 螢-螢、12. 鶯-鶯、13. 錢-錢、14. 賤-賤、
15. 葛 (ヒ) - 葛 (L + 人)、16. 頬-■、17. 祇-■、18. 八-■、19. 北-■、
20. 杖-■、21. 森鷗外-森■外、22. 薩摩-■摩、23. 帝塚山-帝■山、
24. 小樽-小■、25. 剥製-■製

「■」表示のある異体字ペアは以下の通り。ただし 15. は図-1 を参照のこと。

- | | |
|----------------|----------------|
| <p>16. 頬 頬</p> | <p>17. 祇 祇</p> |
| <p>18. 八 八</p> | <p>19. 北 北</p> |
| <p>20. 杖 杖</p> | <p>21. 鷗 鷗</p> |
| <p>22. 薩 薩</p> | <p>23. 塚 塚</p> |
| <p>24. 樽 樽</p> | <p>25. 剥 剥</p> |

4. まとめ

2007年1月以降はJIS X 0213:2004の字体として「葛(L+人)」が多くのIT機器に搭載され、さらに新常用漢字表にも「葛(L+人)」の字体が採用されたため、日本中に広く浸透しつつあると考えられる。本稿を執筆した2012年現在では、山の手住民にも「葛(L+人)」を好む人が増加している可能性が大きい。葛飾・江戸川周辺の住民だけではなく、山の手住民を含む日本全国に「葛(ヒ)」に替わって「葛(L+人)」がなじみ深いものになっていると予想されるからである。

異体字を刺激材料とすれば、文字数、読み、意味が等価で、形や画数だけが異なる刺激ペアを比較的簡単に豊富に準備できる。英語等の諸外国の言語でも、文字数、読み、意味が等しいスペルの刺激ペア(例:center-centre)は存在する。また、「English-eNGLISH」というように、文字数、読み、意味が等価な言語刺激のペアを作成することは不可能ではない。しかし、ペアの一方は現実の生活にはほとんど登場しない、なじみのうすい変異(variation)になるケースが多く、いささか不自然である。その点で、日本語における異体字は国民各層の言語生活に深く浸透しており、日常場面でペアの両方を目にする機会があるものが多いので、適切な刺激だと言えよう。

社会言語学における多変量解析の利用は、世界的に2つの流れがあるように見える。米国では1960年代からLabovがロジスティック回帰分析を積極的に利用してきた実績がある。同じ時期に、日本では「林の数量化理論Ⅲ類」が使われ始めた。米国は仮説検証型の手法であり、日本は仮説探索型の手法である。理想的な研究の進め方は、仮説探索型と仮説検証型の手法をバランスよくミックスし、車の両輪として動かすことであろう。

社会言語学はもとより、心理学などの諸分野を見渡しても、字体の好みから回答者の居住地を推測しようという試みはこれまでなかった。そもそも、そのような発想がなかったと言えよう。日本の社会言語学が、記述的な分析手法のほかに確率モデルに立脚した新たな方法を手にすれば、仮説検証型の研究がさらに盛んになるであろう。言語変化の動向を数量的に予測する研究に向けて一歩踏み出すために、「確率予測の道具立て」を積極的に取り入れる時期が到来したのではないかと考える。

ロジスティック回帰分析は、医学やIT技術などで近年きわめて盛んに利用されるようになった。その背景には、経済学におけるマルコフ連鎖モンテカルロ法(MCMC)や、遺伝子工学のDNA解析におけるベイズ流確率理論の革命的な進歩があるようだが、その説明は紙幅の関係もあるので別の機会に譲りたい。

【附記】

異体字選好データの収集にあたって、笹原宏之氏(早稲田大学教授)、當山日出夫氏(立命館大学非常勤講師)、高田智和氏(国立国語研究所准教授)にご協力をいただいた。また、本稿執筆の準備段階で真田治子氏(立正大学教授)から貴重なコメントを賜った。記して感謝申し上げる。

引用文献

Elliot, R., and Dolan, R. (1998). Neural response during preference and memory judgments for subliminally presented stimuli: A functional neuroimaging study. *The Journal of Neuroscience* 18, 4697-4704.

- Hibiya, J. (1988). A quantitative study of Tokyo Japanese. Doctoral Dissertation, University of Pennsylvania, 1988.
- Kunst-Wilson, W. R., and Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, **207**, 557-558.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Matsuda, K. (1993). Dissecting analogical leveling quantitatively: The case of the innovative potential suffix in Tokyo Japanese. *Language Variation and Change*, **5**, 1-34.
- 笹原宏之・横山詔一・エリク＝ロング (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 (2006a). 「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」『計量国語学』, **25** (5), 199-214.
- 横山詔一 (2006b). 「潜在記憶と言語習得」『月刊言語』, **35**, 52-57, 大修館書店
- 横山詔一・渡邊正孝 (2007). 『記憶・思考・脳』, 新曜社
- Yokoyama, S., and Wada, Y. (2006). A logistic regression model of variant preference in Japanese kanji: an integration of mere exposure effect and the generalized matching law. *Glottometrics*, **12**, 63 -74.
- 横山詔一・真田治子 (2010). 「言語の生涯習得モデルによる共通語化予測」『日本語の研究』, **6** (2), 31-45.
- 横山詔一・笹原宏之・當山日出夫 (2006). 「文字コミュニケーションにおける異体字の選好と親近度：再検査法による信頼性の検討」『社会言語科学』, **9** (1), 16-26.
- 横山詔一・高田智和・當山日出夫・米田純子 (2009). 「平成明朝体のデザイン変異に関する共時的認知分析」『情報処理学会研究報告：人文科学とコンピュータ研究会報告』, 2009 (4), 33-40.
- Zajonc, R.B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, **9**, 1-27.