

台湾日本語学習者は日本人の字体選好をいかに推論するのか

横山詔一*・當山日出夫**・高田智和*・米田純子*

*国立国語研究所, **立命館大学GCOE

キーワード：台湾日本語学習者, 異体字, 選好判断, 推論

日本語の漢字には“異体字 (variant)”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。異体字を刺激材料とすれば、文字数、読み、意味がまったく等価で、形だけが異なる刺激ペアを作成できる。本研究は、台湾の日本語学習者がワープロや携帯メールなどの情報通信機器 (IT 機器) を使って日本人に文章を書くとき、変換候補として「桧一檜」などの異体字ペアが示されると、どちらの字体を選択するのか、その判断は日本人の傾向とどの程度一致するのか、という点について調査・検討した。

Preference for kanji variants in L1 and L2: Predictability based on frequency data

YOKOYAMA Shoichi*, TOYAMA Hideo**, TAKADA Tomokazu*, & YONEDA Junko*

*National Institute for Japanese Language, **Ritsumeikan University

Key words: Kanji variant, preference, implicit inference, politeness, language variation

Japanese kanji characters exhibit variants, which can often be represented by a pair of a traditional and a simplified variants, sharing the same meaning and pronunciation but distinct in visual forms. Studies have shown that preference for variants of competent native readers is explained by relative frequencies of the variant pairs (Yokoyama, 2006). This study investigated the preference for paired variants by second language (L2) readers of Japanese with well-established kanji competencies in their first language (L1) and presumably with more experience with the traditional variants than the simplified counterparts in their L1 orthography. The task for the participants was to choose either one of the paired variants, supposing that they are type-writing letters in Japanese, i.e. the L2, addressed to L1 readers of Japanese, which presumably directed their attention to L2 norms. The participants were junior and senior students, most of whom majored in Japanese at a college in Taiwan. The data showed both similarity and differences between the participants and L1 readers, possibly reflecting the kanji experience of the participants both in L1 and L2. The simplified variants were generally more preferred, as was found in L1 studies. These characters were mostly *joyo* characters, to which the participants were exposed frequently with explicit instruction in the L2.

1. はじめに

本研究は、現代社会における文字コミュニケーションのメカニズムを解明するため、異体字 (variant) に着目した国際比較実験を行った。

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) の豊富さはアジア諸国の漢字文化圏において、日常の文字生活に

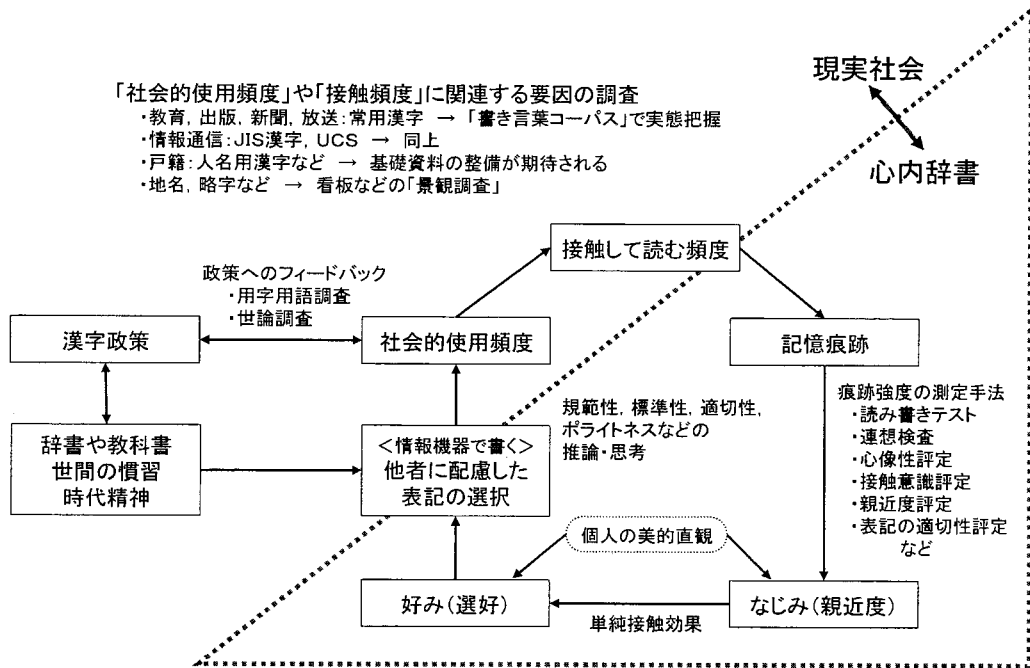


図1 文字生活のサイクルモデル(横山, 2007より)

なんらかの影響をもたらしていると考えられるが、その実態は霧の中にある。そこで、海外の日本語学習者がワープロや携帯メールなどの情報通信機器を使って日本人に文章を書くとき、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選択するのか、その判断は日本人の傾向とどの程度一致するのか、という点について実験的な方法を用いて検討することにした。なお、本稿はJIS X0208をベースに作成しているため、後述の結果の表には印字できない字体が含まれていることをあらかじめ明記しておく。

調査参加者は台湾の大学における日本語学習者であった。台湾と日本は、文字環境がそもそも大きく違う。台湾は繁体字が主流で、日本の常用漢字とは字体が異なる。日本では、日常的に使用する文字の字体基準として、常用漢字

(約2000字)の範囲外に、印刷標準字体(約1000字)が定められている。これは、いわゆる「旧字体」に準拠したものであり、台湾で使用する繁体字と近似する字体が多い。このように、台湾と日本の文字環境は重なる部分と、異なる部分を持っている。

以下、刺激に対して人間が持つ好みを「選好」(preference)、なじみを「親近度」(familiarity)という。選好は書くことに影響する要因で、親近度は読むことに関係する要因だと考えて大過ないだろう。場面設定で「ワープロや携帯メールなどで日本人に文章を書く」という教示を与えた理由は、図1のモデル(横山, 2007)に示す「他者に配慮した表記の選択」を調べるためであった。文章の読み手が日本人であることを念頭におけば、日本人が読みやすい表記、あるいは違和感を持たない表記を意識的・無意識的

に選択するのではないかと考えられる。その際に、日本語としての表記の規範性、標準性、適切性などを台湾日本語学習者は暗黙のうちに推論する必要があるだろう。

より直接的に「日本人はどちらの字体を好むと思いますか」と質問する方法も考えられるが、それは調査参加者の注意をあからさまに字体選好の側面に集中させることになり、日常の文字生活ではほとんど経験したことのない場面を無理にイメージさせる状況になりかねないと判断した。本研究は、なるべく自然な状況を設定し、暗黙のうちになされる「読み手の字体選好」に関する推論（以下、「字体選好暗黙推論」という）にアプローチすることを目指した。

ちなみに、IT 機器を用いて書く、という場面を設定した場合、台湾と日本ではコンピュータ環境が異なるために、異体字ペアが発生する場合と、しない場合がある。今回はこの点についての検討は行わなかった。また、図 1 の文字生活のサイクルモデルが台湾や中国（大陸側）、あるいは韓国で当てはまるかどうかの検討も諸般の事情で割愛した。

1. 1. 字体選好暗黙推論について

台湾から日本に留学している大学生を対象にした調査によると、ワープロや携帯メールなどで文字を書く際により好む字体は旧字体のようである（笹原・横山、私信）。旧字体は新字体よりも繁体字に近い形状を持つため、台湾からの留学生にはある程度の「なじみ」を感じさせるのであろう。つまり、台湾日本語学習者における字体選好は、そもそものベースラインでは、旧字体を新字体よりも高い確率で選択する傾向を有していると考えられる。

台湾日本語学習者に「ワープロや携帯メールなどで日本人に文章を書く」という場面を設定して、字体選好の傾向に変化が生じたとすれば、場面設定による効果があったと判断してよい。その場合の字体選好の変化のパターンは、おおまかに以下の 2 つが予想される。

- 一方が常用漢字である異体字ペアの選好は日本人と同じようなパターンになる。すなわち常用漢字（新字体）が選択される。
- 常用漢字を含まない異体字ペアでは、日本人よりも「新字体」を選好する傾向が強く示される。

仮に以上のような結果が得られたとすれば、台湾日本語学習者が母語の読み書きで選択する繁体字の字体とは違う日本の新字体を選択したことになる。その最大の原因は、「他者に配慮した表記の選択」に帰属できるであろう。読み手として日本人をイメージした場面で、書き手の台湾日本語学習者が自らの判断・意思決定のたよりとするのは、日本語教科書に頻出する字体の記憶像、つまり常用漢字の字体イメージだろうと容易に想像できる。よって、一方が常用漢字である異体字ペアの選好は日本人と同じようなパターンになる。

さらに、台湾日本語学習者は「常用漢字→新字体」という推論を行い、そのルールをほぼすべての異体字ペアに適用すると予想される。その結果、日本人が旧字体を好む異体字ペアについても、台湾日本語学習者は新字体を選ぶことになると考えられる。これは、「過剰適用」や“hyper correction”あるいは方言学研究における“accommodation”の一種と言えよう（阿部、私信）。

場面差が言葉遣いを変化させる研究は話し言葉を中心に行われてきた。しかし、字体選択行動について、場面差の観点から検討を行った研究は寡聞にして知らない。

1. 2. 異体字の頻度調査

異体字は社会でどのくらい読まれているのだろうか。図 1 に示すように、異体字の社会的使用頻度が明らかになれば、人々がどの異体字をどのくらい目にしているかという接触頻度（exposure frequency）を高い精度で推定できるようになる。そのためには、新聞・書籍・看

板・チラシなど、さまざまな活字メディアの資料を大量に収集し、異体字の頻度・用法に関する綿密な調査研究を行う必要がある。

1990年代に新聞コーパスのCD-ROMが市販されるようになると、大量の電子化テキストを利用した異体字研究が始まった。横山・笹原・野崎・ロング（1998）は、新聞コーパスCD-HIASK'93（朝日新聞社、1994：1993年の朝日新聞記事約11万件を収めている）を対象にJIS漢字に含まれる異体字ペア（例：円－圓、万－萬、桧－檜、鶯－鶯、灌－灌など数百字種）の頻度表を示した。笹原・横山・ロング（2003）は、この調査をいわゆる「朝日字体」数千字種に範囲を広げ、頻度と用例の分析を行った。雑誌については1994年度に刊行された雑誌70種類を対象にした200万字規模の漢字頻度表が『現代雑誌の漢字調査』（国立国語研究所、2002）として公刊されている。

さらに、国立国語研究所が現在構築を進めている1億語規模の「書き言葉コーパスKOTONOHA」が完成すれば、新聞・書籍など公共メディアを対象にした異体字研究はより充実するものと期待される。

1. 3. 景観調査や読み書き調査の意義

看板やチラシに使用された漢字の頻度・用法に関しては、最近になって意欲的な研究が増えつつあるように見える。京都「祇園」地区における「祇」の字体を悉皆的に調査している當山（2006、2007）、東京のJR山手線28駅周辺における多言語表示の実態を調査したBackhouse（2005、2006）、東京圏のデパートを対象とした井上（2001）や田中・上倉・秋山・須戸（2007）が代表的な研究である。

景観は公共的な側面をもっているが、今後は調査範囲を公共メディアに限定しないで広く文字生活全般をとらえる努力が求められるであろう。一般人にとって、公共メディアは情報をもつばら「読むこと（reading：受信）の手段」であり、「書くこと（writing：発信）の手段」

ではない。IT機器が爆発的に普及し電子的に文字を書く機会が急速に増大した現在において、若年層の日常生活におけるコミュニケーションは携帯メールなどの個人的なメディア、いわば私的ITメディアを介して盛んに行われている。私的ITメディアは書くだけでなく、読むための手段でもあり、若年層にとっては読み書きの両面に密着したジャンルになっている。そのため、私的ITメディアを射程に入れた調査を行い、公共メディア調査の結果と総合したならば、文字生活における読み書きの実態が浮き彫りになると期待される。

この方面の先駆的研究として『戦後の国民各層の文字生活』（国立国語研究所、1966）があるが、当時はIT環境が整備されていなかったため現代社会の文字生活について直接的な参考とするのは難しい面がある。

1. 4. 実験的手法の長所

私的ITメディアにおける表記行動を範囲に含めた調査を実施しようとする、次の2つのやっかいな問題に直面する（横山・笹原・當山、2006）。

（1）プライバシー保護の問題

私的メディアの文字資料収集には、いつもプライバシー保護や著作権という問題が付きまとう。そのため、私的な文書を多数から提供してもらおうといったことは、実際には非常に難しい。この種の文書は、収集方法によっては、提供者が社会的望ましさなどを暗黙のうちに考慮したフィルターを通過したものとみなすべきものとなり、かなりのバイアスを含んだ表面的な資料だと言わざるをえないだろう。

（2）異体字表示の問題

そもそも文字を書く道具として何を選択するかによって、IT機器では読み書きできない異体字がある。通常の電子メールの場合は、JIS漢字に含まれない字体（字形）は扱えないのが普通である。図2の「6」に掲出した「秤」はどちらか一方の字体しか表示できないのが普通

である。また、たとえば「檜」を使いたいという意向があったとしても、それが JIS 第 2 水準であるために JIS 第 1 水準の「桧」の方が最初の変換候補として表示され、そこで妥協してしまったというようなケースがあるかもしれない。読み書きの道具である IT 機器において発生する「異体字の利用制限」がいかなる影響を及ぼすのか、その影響の大きさをどのように見積もるか、といった点について慎重に検討すべきである。

(3) 実験の手法の導入

以上の理由から、私的 IT メディアを対象にした漢字調査には膨大なコストがかかると見込まれ、将来においても本格的な研究が行われる可能性は低いと考えざるを得ない。

そこで、笹原・横山 (1998) は新たなパラダイムの必要性を唱え、実験的な手法を異体字研究に取り入れた。その手法は簡単で、単漢字の異体字ペアを被験者に呈示して「字体選好課題」を実施するというものである。字体選好課題とは「桧—檜」など 263 字種の新旧字体ペアを被験者に呈示し、IT 機器で字を書く場面をイメージしたとき、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) で直観的に選ばせる課題である。

上記の手法はプライバシー保護や著作権の問題とは無関係であるし、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・印字できるようフォントを準備して紙ベースで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もない。

2. 方法

2. 1. 刺激材料について

刺激材料とした異体字ペアの一部の例を下図 2 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た。刺激項目は、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 2005 年にデザインした約 6 万字の平成明朝体

グリフから選んだ新旧 120 ペアであった。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 120 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。

3	壺	壺
6	秤	秤
11	蚩	螢
14	鶯	鶯
17	会	會
18	檜	桧
21	葛	葛
22	喝	喝
23	觀	觀
24	灌	灌

図 2 異体字ペアの例

上記の規準に適合した異体字集合から、さらに、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、平成明朝グリフを 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、学籍番号、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。

2. 2. 実験参加者

台湾日本語学習者のデータは台北市内にある東呉大学日本語学科の3年生と4年生のほか日本語を副専攻とする学生計78名から収集した。年齢は20歳から24歳が大部分を占めていたが、25歳と28歳が1名ずつ含まれていた。調査は2007年5月に東呉大学において実施された。

日本語母語話者のデータは京都市内にある立命館大学の学生150名(18歳から23歳)から収集した。調査は2006年1月に立命館大学において実施された。データのうち未記入・無回答は欠損値として扱い、除外した。よって字体ペアごとに標本数は多少異なっていた。

2. 3. 台湾日本語学習者による字体選好暗黙推論の手続き

(1) 教示

台湾日本語学習者に異体字のペアを示して、より使いたいと感じる方の字を選択させた。調査における教示は次のようなものであった。「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。たとえば日本語では、「断」と「斷」は同じ読みで同じ意味の漢字のペアです。もし、あなたがパソコンで日本語の文章を書いて日本人に送るとしたら、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)

(2) 書記経済の配慮

文字を自らの手で書く場面では簡単な字体を選ぶ人が多くなるという可能性もあるため、パソコン画面上で文字を選択する場面をイメージさせてその影響を小さくするよう心がけた。

(3) 表記に気を遣う相手や場面の設定

この課題は、日本人に向けて文章を書くという場面を設定した。これは、図1の「他者に配慮した表記の選択」を支える「暗黙の推論」を調べるための工夫であった。

2. 4. 日本語母語話者による字体選好の手続き

異体字のペアを示して、より使いたいと感じる方の字を選択させた。調査における教示は以下の通り。「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、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。(以下略)

3. 結果

異体字ペアごとに旧字体を選択した人数のパーセントを求め、旧字体選好率とした。以下の数表には旧字体選好率を示す。

3. 1. 台湾日本語学習者と日本語母語話者で差が大きい字体

(1) 新字体に偏るケース

台湾日本語学習者が日本人に向けて文章を書くときの旧字体選好率が、日本人の旧字体選

好率を30%以上も下回った異体字ペアを表3.1.に示す(異体字ペアの一方が掲出されていない部分があるが、これはパソコンにフォントが搭載されていないためである)。ここには日本の常用漢字はほとんど含まれていないことが分かる。異体字ペアの字体差をみると、新旧間で違いが大きいものが多い。

表 3.1. 字体選好暗黙推論が新字体に偏る例

ID	異体字ペア	常用など	台湾 (%)	日本 (%)	台湾から日本を減じた差
3	壺壺		21.8	90.7	-68.9
24	灌灌		19.2	86.0	-66.8
176	籠籠		14.1	76.7	-62.6
174	竜龍	△	7.7	70.0	-62.3
14	鶯鶯		10.3	62.7	-52.4
122	賤賤		17.9	63.3	-45.4
138	鋳鋳		17.9	63.3	-45.4
37	鴟鴟		12.8	55.3	-42.5
175	滝瀧	○	2.6	44.0	-41.4
28	俠俠		10.3	50.0	-39.7
18	桧檜		38.5	75.3	-36.8
228	囊囊		11.5	48.0	-36.5
246	撥撥		13.0	49.3	-36.3
43	鱒鱒		5.1	38.7	-33.6
141	蟬蟬		7.7	41.3	-33.6
180	諫諫		17.9	49.3	-31.4
156	澆澆		14.1	45.3	-31.2
166	蛎蠣		9.0	40.0	-31.0

(2) 旧字体に偏るケース

台湾日本語学習者が日本人に向けて文章を書くときの旧字体選好率が、日本人の旧字体選好率を30%以上も上回った異体字ペアを表3.2.に示す(異体字ペアの一方が掲出されていない部分があるが、これはパソコンにフォントが搭載されていないためである)。

「飲」を除くと、日本の常用漢字は含まれていない。異体字ペアの字体差をみると、新旧間

での違いはあまり大きくないものが多い。たとえば、点の方向や筆法差といったレベルの微差にとどまるペアが少なくないように見える。字体間の違いが微差の場合は、普段の生活でよく接触している「なじみ」のある字体を無意識のうちに自動的に選択する傾向があるのかもしれない。

表 3.2. 字体選好暗黙推論が旧字体に偏る例

ID	異体字ペア	常用など	台湾 (%)	日本 (%)	台湾から日本を減じた差
99	蝕蝕		47.4	16.0	31.4
88	哨哨	新	56.4	24.0	32.4
72	溺溺		64.1	30.0	34.1
158	稗稗		61.5	26.7	34.8
220	蔑蔑		74.4	38.7	35.7
89	屑屑		55.1	17.3	37.8
6	秤秤		65.4	26.7	38.7
73	鱒鱒		66.7	24.0	42.7
96	飲飲	○	44.9	1.3	43.6

3. 2. 台湾日本語学習者と日本語母語話者で差が小さい字体

台湾日本語学習者が日本人に向けて文章を書くときの旧字体選好率が、日本人の旧字体選好率と差が10%以内のものを表3.3.に示す。ここには日本の常用漢字が多く含まれている。

表 3.3. 字体選好暗黙推論が的中した例

ID	異体字ペア	常用など	台湾 (%)	日本 (%)	台湾から日本を減じた差
125	曾曾	人	20.5	29.3	-8.8
164	万萬	○	1.3	10.0	-8.7
157	醜醜		9.1	17.3	-8.2
54	国國	○	2.6	10.7	-8.1
29	狹狹	○	5.1	11.3	-6.2
38	経経	○	1.3	6.7	-5.4
222	隋隋		16.7	22.0	-5.3

136	駅驛	○	5.1	10.0	-4.9
171	瑤瑤	人	15.4	20	-4.6
150	読讀	○	2.6	6.7	-4.1
23	観観	○	1.3	4.0	-2.7
242	琢	人	32.1	34.7	-2.6
74	寿壽	○	7.7	10.0	-2.3
155	発發	○	2.6	4.7	-2.1
17	会會	○	1.3	3.3	-2.0
104	楨楨	人	29.5	31.3	-1.8
44	儉儉	○	5.1	6.7	-1.6
139	単單	○	3.9	5.3	-1.4
137	訳譯	○	5.1	5.3	-0.2
108	尽盡	○	5.1	4.7	0.4
34	区區	○	5.1	4.0	1.1
95	嬢嬢	○	10.3	8.7	1.6
121	錢錢	○	6.4	4.7	1.7
110	数數	○	3.8	2.0	1.8
133	啄	人	34.6	32.0	2.6
148	売賣	○	6.4	3.3	3.1
170	謡謡	○	15.4	12.0	3.4
102	填		28.2	24.7	3.5
140	戦戦	○	7.7	3.3	4.4
63	祇		70.5	64.0	6.5
41	倦		29.5	22.7	6.8
178	蠟		39.1	32.0	7.1
126	噌		21.8	14.0	7.8
42	捲		26.9	18.0	8.9

4. 考察

以上の結果から、一方が常用漢字である異体字ペアに対する台湾日本語学習者の選好は日本人と同じようなパターンになり、常用漢字（新字体）が選択される確率が高いということが確認された。また、常用漢字を含まない異体字ペアでは、日本人よりも「新字体」を選好する傾向が強く示されることも分かった。ただし、そのような傾向は異体字ペアで新旧字体差が大きい場合に多くみられることが明らかになった。さらに、常用漢字を含まない異体字ペアで字体

差が点の方向や筆法の差といった微差にとどまる場合は、台湾日本語学習者の選好は日本人よりも「旧字体」に偏る傾向が認められた。これらは、先に述べた「他者に配慮した表記の選択」という要因で説明できると考えられる。あらかじめ予測した結果のパターンが得られたと言って大過ない。

今後の検討課題を4つ述べる。第1に、台湾日本語学習者における字体選好は、そもそものベースラインでは、旧字体を新字体よりも高い確率で選択する傾向を有しているという、より具体的な証拠が必要である。本研究はその点が脆弱である。第2に、日本語学習の教科書で、実際に使用されている漢字はどうか、調査すべきである。その延長には、台湾に流入している、サブカルチャーとしての種々の日本文化（アニメなど）も視野にいれる必要があるだろう。第3に、拡張新字体や「澤」など、個別に検討・吟味すべき文字が少なからずあるように思える。第4に、日本のパソコンの文字と、台湾のパソコンでは、そもそも搭載字数・字体が異なる。この影響をどう考えるべきか。

いずれにせよ、本研究は、日本語だけではわからない、文字についての様々な事象を考察する手がかりを手得るための第一段階にはなったと考える。今後は日本語能力のレベルについてのより綿密なデータもあわせて採取するなどの工夫が必要になるだろう。

【謝辞】

台湾で調査を実施するにあたり、台北市にある東呉大学の田中章夫教授と陳淑娟教授、台湾大学の林立萍副教授、ならびに日本の埼玉学園大学の真田治子准教授から多大なるご協力とご支援を賜りました。また、早稲田大学の笹原宏之教授には本研究全般について貴重なご助言をいただきました。英文概要作成は国立国語研究所非常勤研究員の和田志子氏から支援を受けました。ここに記して感謝申し上げます。

主要な参考文献

- 国立国語研究所 (1966) 『戦後の国民各層の文字生活』 (国立国語研究所報告 29), 秀英出版
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980) Affective discrimination of stimuli that cannot be recognized. *Science*, 207, pp.557-558.