北京大学学报(自然科学版)
Acta Scientiarum Naturalium Universitatis Pekinensis
doi: 10.13209/j.0479-8023.2015.043

# 语料库语言学视角下的台湾汉字简化研究

王博立　史晓东†　陈毅东　任文瑶　阎思瑶

厦门大学智能科学与技术系，厦门　361005; † 通信作者, E-mail: mandel@xmu.edu.cn

**摘要**　采用语料库语言学的研究方法，论证了台湾存在汉字简化的现象，并分析了台湾汉字简化的特点和影响因素。首先，通过爬取台湾的新闻媒体、政府网站和博客，建立台湾汉语语料库。然后，借助语料库统计数据论证了台湾民间存在对简笔俗字的使用偏好，即台湾存在汉字简化的现象，并进一步分析台湾汉字简化的若干特点。最后，讨论了台湾汉字简化的影响因素，包括大陆简体字、汉字编码、中文输入法等。

**关键词**　语料库语言学；台湾；汉字；汉字简化；俗字

**中图分类号**　TP391

# On the Simplification of Chinese Characters in Taiwan: A Perspective of Corpus Linguistics

WANG Boli, SHI Xiaodong†, CHEN Yidong, REN Wenyao, YAN Siyao

Cognitive Science Department, Xiamen University, Xiamen 361005; † Corresponding author, E-mail: mandel@xmu.edu.cn

**Abstract**　Corpus linguistics methods were applied to prove that there is simplification phenomenon of Chinese characters in Taiwan. Firstly, a Taiwan Chinese corpus was built up with a large number of texts from media, government website and blog. Secondly, with statistics from corpus, it was proved that civilians in Taiwan prefer to use those popular Chinese characters with fewer strokes, which implies a simplification phenomenon. Lastly, the authors analyzed several influential factors of the simplification of Chinese characters in Taiwan, including simplified Chinese from mainland, Chinese character encoding and Chinese IME.

**Key words**　corpus linguistics; Taiwan; Chinese characters; simplification of Chinese characters; popular Chinese characters

由于政治上的对立和分隔，海峡两岸的文化交流在 20 世纪 80 年代之前几乎完全停滞，致使两岸在语言文字的使用习惯上形成一定的差异，表现在字音、拼读系统、标点符号、书写系统、词汇、语法以及中文排写等多个方面，其中以书写系统的差异最为明显。两岸书写系统的差异常被概括地认为是汉字字形上简体字与繁体字(或台湾所言“正体字”)的差异。但本研究发现，实际上两岸字形方面的差异十分复杂，台湾亦存在汉字字形简化的现象。我们收集了大规模的台湾语料，建立台湾汉语语料库，借助该语料库，采用定性与定量相结合的研究方法，研究台湾的汉字简化现象及其影响因素。

## 1　相关概念界定

本节对研究台湾汉字所涉及的简体字、繁体字、俗字、汉字简化等概念做出界定，以便于下文讨论。

1) 简体字：指大陆推行《简化字总表》后的中文书写系统。

教育部专项“简繁汉字智能转换系统”，国家科技支撑计划项目(2012BAH14F03)，国家自然科学基金(61303082，61005052)，教育部博士点基金(20130121110040)资助

收稿日期：2014–06–30；修回日期：2014–11–09；网络出版时间：2014-11-28 14:21

2) 繁体字: 与“简体字”相对, 指由历史上流传下来、目前仍在台湾香港等地广泛使用的传统汉字①。

3) 俗字: 亦称“俗体字”、“俗写”、“简写字”、“手写简笔字”、“手头字”、“破体字”、“小写”等等, 指流行于民间, 有别于官方认定的“正体字”的另一种字体[1], 可以认为俗字是在民间约定俗成广泛使用的异体字, 且通常具有较为简单的字形。由于俗字的这两个特点, 在汉字简化过程中, 往往采用俗字作为简体字形, 是大陆简体字的重要来源。台湾“教育部”于 1979 年公布了《标准行书范本》, 梳理了台湾民众习惯使用的简笔俗字。

4) 汉字简化: 指在汉字的实际使用中, 逐渐以笔画较简的字代替笔画较繁的字, 即在整个社会范围内, 一部分繁体字的使用频率逐渐降低, 而那部分与之对应的笔画较简的字的使用频率逐渐提高。

## 2 语料库建设

我们从互联网上收集、爬取, 并加以整理, 得到一个规模为 17 亿字的台湾汉语语料库。目前该语料库已经在互联网上公开②, 并且规模仍然在不断扩充。如表 1 所示, 该语料库依据来源划分为 8 个子语料库, 语料内容涵盖政府公文、新闻和博客 3 种不同类型的文本, 语料的时间跨度为 1991 年至今。本研究主要使用发布时间早于 2013 年 12 月 31 日的语料。

值得注意的是, 语料中夹杂着少数用字错误: 一些媒体或博主所发布的内容系原始简体文本经低质量的简繁自动转换系统转换后得到的; 在博客语料中甚至还存在着一定数量的大陆简体文本。

## 3 台湾汉字简化的表现与特点

与大陆行政指令驱动下的汉字简化运动不同, 台湾的汉字简化主要表现为, 在汉字的实际使用过程中大量使用民间俗字。本节利用语料库资源, 采用统计方法证明台湾存在汉字简化的现象, 并讨论台湾汉字简化的特点。

**表 1 台湾汉语语料库来源分布情况**

Table 1 Distribution of Taiwan Chinese corpus source

| 名称 | 规模/万字 | 年代 | 类型 |
|---|---|---|---|
| CNA 新闻 | 10000 | 2007—2013 | 新闻 |
| gigaword CNA | 76000 | 1991—2004 | 新闻 |
| 台湾政教 | 4213 | 不详—2013 | 政府公文、新闻 |
| 苹果日报 | 4534 | 2011—2013 | 新闻 |
| 人间福报 | 15000 | 2000—2013 | 新闻 |
| 台湾 msn | 16000 | 2011—2014 | 新闻 |
| 台湾 yahoo | 14000 | 2009—2014 | 新闻 |
| 无名小站部落格 | 33000 | 不详—2013 | 博客 |

### 3.1 民间俗字的大量使用

语料库的统计数据显示, 台湾的民间俗字被大量使用。相比于“国字标准字体”③中收录的正体字, 或台湾“教育部”《重编国语辞典修订本》④中规定的规范用字, 台湾媒体在一些情况下更喜欢采用民间的简笔俗字, 尽管这些俗字只有异体字的地位。

表 2 列举了几个俗字在新闻媒体中的使用情况。其中, “台”本身是正字, 但其作为正字的用法在现代文中极为少见⑤, 大部分情况下, “台”字以“臺”的异体字的身份出现(如“台灣”“台北”), 而且从表 2 中“吧台”一项可以看出, 媒体中也出现了将“台”用作“檯”的简体字的情况⑥。《标准行书范本》中, “銹”是“鏽”的手写简笔字, 《重编国语辞典修订本》仅收录“鏽”而无“銹”, 而台湾“教育部”《异体字字典》⑦认为“銹”为“鏽”之异体。但从表 2 可以看出, “銹”字在媒体中有相当频率的使用。类似的, “厘”是《标准行书范本》中“釐”的手写简笔字, 从表 2 可以看出, “厘”字在媒体中也有相当频率的使用(特别是用作“公厘”时)。

由上述分析可知, 台湾媒体在用字上存在舍弃繁难正体字而使用简笔俗字的现象。这种对简笔俗字的使用偏好正是台湾汉字简化的表现。与大陆行政指令自上而下的汉字简化不同, 台湾这种源于简

① 在台湾地区, 有时会使用“正体字”或“正字”与大陆“简体字”相对应, 等同于“繁体字”。

② 该语料库可通过 http://corpus.superfection.com/corpus_tc.html 进行检索。

③ “国字标准字体”主要包括《常用国字标准字体表》和《次常用国字标准字体表》。

④ 台湾“教育部”《重编国语辞典修订本》http://dict.revised.moe.edu.tw/。

⑤根据《重编国语辞典修订本》, “台”作正字时用于“台州”、“天台山”等地名, 或在古文中用作“怡”或“我”之义, 或表疑问。

⑥ 《标准行书范本》中, “檯”的手写简笔字应为“枱”。

⑦ 台湾“教育部”《异体字字典》http://dict2.variants.moe.edu.tw/variants/。

**表 2　部分简体字在新闻语料中出现的相对频率统计**

Table 2　Relative frequency of several simplified Chinese characters in news corpus

| 简体字 | 所有相关字形 | 语料库中简体字出现的相对频率/% | | |
|---|---|---|---|---|
| | | CNA 新闻 | 苹果日报 | 人间福报 |
| 台 | 台/臺/檯/颱/枱 | 96.734 | 94.452 | 97.539 |
| 國立[台]灣大學 | 國立[台/臺]灣大學 | 96.624 | 76.923 | 90.446 |
| 吧[台] | 吧[台/臺/檯/枱] | 48.571 | 54.386 | 79.817 |
| 銹 | 銹/鏽/锈 | 36.700 | 22.581 | 30.034 |
| 不[銹]鋼 | 不[銹/鏽/锈]鋼 | 28.879 | 37.662 | 33.784 |
| 厘 | 厘/釐 | 3.455 | 7.270 | 24.145 |
| 公[厘] | 公[厘/釐] | 25.279 | 64.609 | 32.115 |

说明: “相对频率”指简体字出现的频次与所有相关字形出现的频次之比。

笔俗字的渐变的汉字简化, 更尊重传统的语言文字和民间的用字习惯, 带有草根特色, 符合语言发展固有的渐变性和规律性, 但正异字并存的情况也造成一些用字不规范的问题。

### 3.2　台湾官方对俗字使用的保守态度

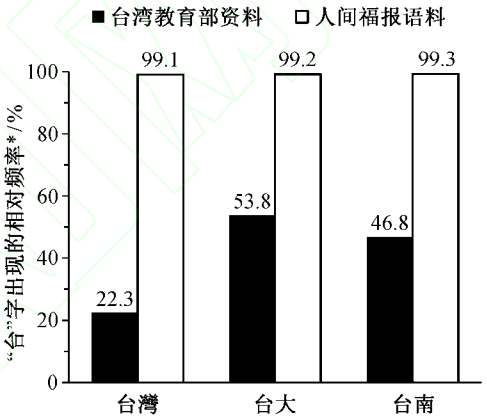
与大陆的汉字简化历程相比, 台湾官方在汉字简化问题上显得比较保守和消极。国民党政府曾于 20 世纪 30 年代和 50 年代两次研究制定汉字简化方案, 但均因政界和学界保守势力的反对, 最终无法实施[2]。前文提及的《标准行书范本》也仅仅是手写行书规范①, 台湾“教育部”并未对其进行大力推广, 对汉字简化的影响不大。

相反的是, 民间对简笔俗字的使用偏好对台湾官方的用字产生了较大影响。图 1 对比了台湾“教育部”语料②和人间福报语料上“台灣”、“台大”、“台南” 3 个词语中, 简体“台”字出现的相对频率。从图 1 可以看出, 台湾官方也会使用简体“台”来代替正字“臺”, 且有相当比例。但与媒体相比, 台湾官方对简体字的使用明显较为保守, 这很可能与其强调推广“正字”或“国字”的主张有关。

### 3.3　正体字与异体字的混用

简笔俗字的大量使用造成正体字与异体字并存, 进而产生许多用字习惯混乱、用字不规范的现象。例如, “台”为“臺”的简笔俗字, “檯”的简笔俗字为“枱”, 但在表 2 所列举的“吧台”一词中, 出现以“台”代“檯”的现象, 不符合用字规范。

在这样的情况下, 甚至出现了一些正体与异体




“相对频率”指简体字出现的频次与所有相关字形出现的频次之比, 如在“台灣”一词中, “台”字出现的相对频率等于语料中“台灣”的频次除以“台灣”和“臺灣”的频次之和。

**图 1　2012—2013 年语料库中“台”字相对频率统计**

Fig. 1　Relative Frequency of Character Tai in Corpus among 2012–2013

倒置的现象。例如, 根据《常用国字标准字体表》《重编国语辞典修订本》《异体字字典》, “晒”为正体, “曬”为异体, 但从表 3 的统计数据可以看出, 无论是博客(无名小站语料)、媒体, 还是官方(台湾政教语料), 均明显倾向于使用“曬”, 尽管“晒”为正体且字形明显比“曬”简单, 类似的情况还有“癡”与“痴”、“薦”与“荐”等。这种对繁难异体字的使用偏好, 可能有两个方面的原因: 其一是台湾社会对大陆简体字的排斥心理, 其二是中文输入法的选字, 后文将分别对二者进行讨论。

① 印刷品需采用“国字标准字体”。

② 台湾“教育部”语料由台湾“教育部”全球资讯网检索得到(http://www.edu.tw/search.aspx)。

**表 3　语料库中“晒”“曬”二字相对频率统计**

Table 3　Relative frequency of character Shai in corpus

| 语料库 | 相对频率/% | | | | | | |
|---|---|---|---|---|---|---|---|
| | CNA 新闻 | 台湾政教 | 无名小站 | 苹果日报 | 人间福报 | 台湾 msn | 台湾 yahoo |
| 晒 | 2.137 | 30.769 | 5.563 | 1.340 | 7.349 | 3.377 | 2.997 |
| 曬 | 97.863 | 69.231 | 94.437 | 98.660 | 92.651 | 96.623 | 97.003 |

### 3.4　汉字数量繁多

大陆的汉字简化，指导思想是从形体和数量两方面对汉字进行精简[3]，被简化的繁体字不再使用，目前基本实现了这一目标。而台湾的汉字简化，主要是形体上的精简，而并无数量上的精简，大量使用简笔俗字的同时，也仍然使用原来的繁体字。表 4 的统计数据显示，在大陆的新华网语料①中，字频最高的 3888 个汉字能覆盖 99.9%的新闻文本，而相同的覆盖率在台湾 msn 语料中需要字频最高的 4101 个汉字；大陆新华网语料中字频最高的 100 个汉字能覆盖 37.99%的新闻文本，而台湾 msn 语料中字频最高的 100 个汉字只能覆盖 35.33%的新闻文本。这表明，当表示相同总量的信息时，台湾人需要使用比大陆人更多的汉字，说明繁体汉字对于信息的传播和知识的普及存在一定的阻碍，简化汉字的确有助于汉语学习，一定程度上减轻了学生的学习负担。

**表 4　两岸新闻语料中汉字覆盖情况对比**

Table 4　Chinese character coverage in Chinese corpus of Mainland and Taiwan

| 语料 | 覆盖率超 99.9%时汉字的个数 | 前 100 个高频汉字的覆盖率/% |
|---|---|---|
| 新华网(10 亿字) | 3888 | 37.99 |
| 台湾 msn (1.3 亿字) | 4101 | 35.33 |

## 4　台湾汉字简化的影响因素

在长期的使用过程中，台湾汉字的字形和使用习惯受到一些外部因素的影响。大陆简体字、汉字编码、中文输入法等因素都在一定程度上影响着台湾的汉字简化。

### 4.1　大陆简体字

随着大陆国际影响力的日益增强和海峡两岸经济文化交流的不断深入，大陆简体字对台湾用字习惯的影响也日益加深。据媒体报道，在台湾的一些旅游景点，为了吸引大陆游客，不少商家在宣传材料上使用简体字[4]。语料库的统计数据也反映了大陆简体字的影响。表 5 列举了一些大陆所用简体字在台湾语料中的出现情况，从一个侧面反映了大陆简体字对台湾用字习惯产生的影响。此外，台湾一些简笔俗字的使用范围有所扩大，与大陆的使用习惯趋同。例如，“台”字本为“臺”的简笔俗字，“檯”的简笔俗字本为“枱”。但表 2 中的“吧台”一项将“台”作为“檯”的简笔字，并且“梳妝檯”“寫字檯”等词也存在类似的情况。台湾这种简笔俗字使用范围的扩大，正是受到大陆用字习惯的影响，因为

**表 5　部分简体字在台湾语料中出现的相对频率统计**

Table 5　Relative frequency of several simplified Chinese characters in Taiwan Chinese corpus

| 简体字 | 所有相关字形 | 语料库中简体字出现的相对频率/% | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | CNA 新闻 | 台湾政教 | 无名小站 | 苹果日报 | 人间福报 | 台湾 msn | 台湾 yahoo | 合计 |
| 污 | 污/汚/汙 | 80.79 | 98.05 | 70.49 | 81.96 | 54.78 | 59.16 | 65.30 | 70.22 |
| 坂 | 坂/阪 | 3.82 | 23.33 | 14.10 | 16.09 | 8.38 | 13.19 | 13.70 | 13.26 |
| 羡 | 羡/羨 | 6.73 | 6.12 | 3.56 | 4.74 | 10.23 | 9.09 | 9.74 | 5.70 |
| 亘 | 亘/亙 | 0 | 28.00 | 21.96 | 96.60 | 2.75 | 7.27 | 11.25 | 28.99 |
| 疱 | 疱/皰 | 41.89 | 72.92 | 15.84 | 60.50 | 27.70 | 49.44 | 51.44 | 37.35 |

说明：表中所选简体字均为大陆《通用规范汉字表》收录而台湾《重编国语辞典修订本》和“国字标准字体”未收录的汉字。

① 该语料库可通过 http://corpus.superfection.com/corpus_sc.html 进行检索。

大陆简化字中将“台”“臺”“檯”“颱”合并为“台”。

对于来自大陆简化字的影响, 台湾社会也出现了一种抵触心理, 认为繁体字(即台湾所称正体字)代表中华传统文化, 应尽量避免使用简体字, 保护和推广繁体字。据报道, 台湾领导人马英九多次强调“正体汉字”是“中华文化的精髓”, 提倡使用正体汉字, “不要为了招揽(大陆)观光客, 而在招牌或文宣上使用简体汉字”[5]。3.3 节提到的台湾出现的个别对繁难异体字的偏好, 很可能是在这种社会心理的驱动下而刻意为之的结果, 认为相较于简笔的正体字, 使用繁难的异体字更能体现传统文化。

## 4.2 汉字编码

在当今信息化时代的背景下, 汉字编码是影响台湾用字习惯的一个重要因素。Big5 码是过去台湾地区最通行的计算机汉字编码方式, 根据《常用国字标准字体表》《次常用国字标准字体表》等汇编而成, 收录汉字 13000 多个, 但未收录被视为异体字的部分民间俗字, 如“着”“塗”“煊”“喆”“銹”等, 导致这些民间常用的俗字无法在计算机中正常显示[6]。

近年来, 不少台湾软件改用字库规模更大的国际标准 Unicode 进行编码, 但其中亦存在不少问题。例如, 图 2 中的汉字在 Unicode 中被认为是同一编码不同字体的差异[7], 而图 3 中的汉字却被认为是不同编码的差异。其实从字形来看, 这两组字的差异都不大, 区分是否应该采用同一编码或不同编码的标准不明确。

造成 Unicode 汉字编码混乱的直接原因是两岸及日、韩等国家均各自向 Unicode 联盟提交汉字编码方案, 合并时存在一些重复[8]。而涉及大陆与台湾的重复编码, 根本原因是两岸所采用的印刷字体①在一些构字部件上存在差异。台湾的“国字标准字体”基本沿用古籍所用“旧字形”; 而大陆在整理汉字时则依据“从俗从简”的原则, 于 1965 年修订《印刷通用汉字字形表》, 称为“新字形”[9]。不少新旧字形的差异被误认为是简繁关系, 错误地将其作为两个汉字来对待, 在 Unicode 中赋予不同的编码。但实际上, 这种字形差异是同一汉字在不同字体下表现出的差异, 应当赋予相同的编码②。


**图 2 Unicode 中部分同码字在两岸具有不同字形的情况**

Fig. 2 Chinese Characters with same Unicode but different structure in mainland and Taiwan

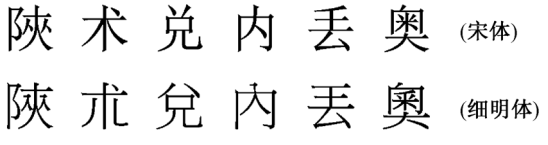



**图 3 Unicode 中部分异码字在两岸具有相近字形的情况**

Fig. 3 Chinese Characters with different Unicode but similar structure in mainland and Taiwan

## 4.3 中文输入法

在信息化高度发达的今天, 中文输入法也是影响台湾汉字简化的一个重要因素。汉字输入法主要包括字形输入法和字音输入法两类③。台湾目前比较流行的输入法中, 仓颉输入法和呒虾米输入法为字形输入法, 而新注音输入法等为字音输入法, 其中新注音输入法的用户量接近五成[10]。

字音输入法中, 同一音码对应的候选词的排列顺序对台湾汉字字形具有一定的影响。在使用这类字音输入法时, 对于一组同音通行词, 用户通常会优先选择排序靠前者(往往是默认首选词)。表 6 为新注音输入法的连打功能对部分台湾通行词的输入情况。在拥有大量用户的情况下, 新注音输入法的选词倾向将直接对台湾互联网的用字习惯产生影响。如表 6 中所示, 使用新注音输入法的用户就会出现“周邊”与“電腦週邊”、“週六”与“周日”倾向使用不同的“周”, 这使得同一意义的用字习惯无规律可循。

## 4.4 其他可能的影响因素

新加坡和马来西亚广泛采用简体汉字。自 1976 年颁布《简体字总表》修订本, 新加坡的简体字与中国大陆完全一致, 但新加坡和马来西亚民间仍广泛使用繁体字[11]。新马两地华人长期以来与台湾保持密切的联系, 是大陆简体字影响台湾的一个重要途径。

① 此处指官方所确定的规范印刷字体。

② 正如同一个汉字在楷体、行书、隶书中, 虽然存在字形差异, 但仍是同一个汉字, 应当采用同一编码。

③ 字形输入法使用汉字的形码进行输入, 大陆使用的五笔输入法和笔画输入法即属于字形输入法; 字音输入法使用汉字的音码进行输入, 大陆使用的拼音输入法即属于字音输入法。

**表 6　新注音输入法连打功能输入字形情况**

Table 6　Input result of New Zhuyin IME using full word input style

| 简体词形 | 台湾通行词 | 默认首选词形 |
|---|---|---|
| 台湾 | 台灣/臺灣 | 台灣 |
| 舞台 | 舞台/舞臺 | 舞台 |
| 周边 | 周邊/週邊 | 周邊 |
| 电脑周边 | 電腦週邊/電腦周邊 | 電腦週邊 |
| 周六 | 周六/週六 | 週六 |
| 周日 | 周日/週日 | 周日 |
| 公布 | 公布/公佈 | 公佈 |
| 制作 | 製作/制作 | 製作 |
| 采访 | 採訪/采訪 | 採訪 |
| 品尝 | 品嚐/品嘗 | 品嚐 |

香港和澳门地区与大陆联系密切，不少大陆人口到港澳地区定居生活，将一些大陆的用字习惯带到港澳地区，造成港澳地区汉字使用较为混乱的情况，如在繁体字文本中掺杂简体字或不规范地使用繁体字。这些不规范的用字情况也随着港澳与台湾的交流影响到台湾的用字情况。此外，香港政府颁布的基于台湾 Big5 码的扩展“香港增补字符集”(HKSCS，含粤语汉字及部分异体字和简体字[12])被不少繁体中文系统采用，对台湾地区也产生一定的影响。台湾民间发起的“Unicode 补完计划”收录了“香港增补字符集”中的一部分汉字[13]。

## 5　结论

通过上述分析，本文得出如下结论。

1) 台湾也存在汉字简化的现象，表现为民间对简笔俗字的使用偏好，与大陆的简化模式截然不同。

2) 台湾官方在汉字简化过程中扮演保守消极的角色，但其用字习惯也受民间俗字影响。

3) 台湾在使用民间俗字的同时也继续使用繁难的正体字，这种正异并存的情况，一方面造成一些用字不规范的问题，另一方面使得台湾的汉字简化并未减少汉字的数量，在一定程度上不利于信息的传播和知识的普及。

4) 在两岸交流不断深入的背景下，大陆简体字对台湾的影响日益显著，这也导致台湾社会对大陆简体字存在一定的抵触心理。

5) 在信息化时代的背景下，汉字编码的方式和字音输入法中候选词的排序对台湾的用字习惯具有一定影响。

## 参考文献


[1] 维基百科. 俗字[DB/OL]. (2013–06–30) [2014–01–21]. http://zh.wikipedia.org/wiki/%E4%BF%97%E9%AB%94%E5%AD%97

[2] 熊皮匠的文字作坊. 民国时及其后的台湾当局，三次试图进行的汉字简化[EB/OL]. (2012–02–27) [2014–01–21]. http://blog.sina.com.cn/s/blog_5188d1380102dtiz.html

[3] 新华网. 书同文:《汉字简化方案》制订始末[N/OL]. (2008–06–03) [2014–01–21]. http://news.xinhuanet.com/theory/2008-06/03/content_8304343.htm

[4] 新华每日电讯. 简体字在台湾：最能吸引大陆游客目光[N/OL]. (2009–07–19) [2014–01–26]. http://news.xinhuanet.com/mrdx/2009-07/19/content_11731409.htm

[5] 中国新闻网. 马英九出席汉字文化节：不应为招揽陆客使用简体字[N/OL]. (2014–01–02) [2014–02–04]. http://www.chinanews.com/tw/2014/01-02/5687464.shtml

[6] 维基百科. 大五码[DB/OL]. (2013–10–22) [2014–01–26]. http://zh.wikipedia.org/wiki/Big5

[7] 中日韩汉字求同存异[DB/OL]. (2013–10–22) [2014–1–26]. http://hanzi.unihan.com.cn/CoolHanzi/

[8] 维基百科. 中日韩越统一表意文字[DB/OL]. (2014–01–24) [2014–01–26]. http://zh.wikipedia.org/wiki/CJK

[9] 维基百科. 新字形[DB/OL]. (2014–01–16) [2014–02–01]. http://zh.wikipedia.org/wiki/%E6%96%B0%E5%AD%97%E5%BD%A2

[10] Pollster 波仕特線上市調. 七成以上民眾使用注音輸入法[R/OL]. (2011–07–07) [2014–01–26]. http://www.pollster.com.tw/Aboutlook/lookview_item.aspx?ms_sn=1476

[11] 维基百科. 新加坡汉字[DB/OL]. (2014–01–04) [2014–02–01]. http://zh.wikipedia.org/wiki/%E6%96%B0%E5%8A%A0%E5%9D%A1%E6%BC%A2%E5%AD%97

[12] 维基百科. 香港增补字符集[DB/OL]. (2013–08–14) [2014–02–01]. http://zh.wikipedia.org/wiki/%E9%A6%99%E6%B8%AF%E5%A2%9E%E8%A3%9C%E5%AD%97%E7%AC%A6%E9%9B%86

[13] 维基百科. Unicode 补完计划[DB/OL]. (2014–03–02) [2014–03–19]. http://zh.wikipedia.org/wiki/Unicode%E8%A3%9C%E5%AE%8C%E8%A8%88%E7%95%AB