

# ISO-10646 国际编码字符集存在的问题

河北大学人文学院 杨宝忠

ISO-10646 国际编码字符集（以下简称“字符集”）的研制，对于汉字的数字化、国际化具有积极意义；但由于字符集追求收字数量而忽视字形整理，致使字符集中存在不少问题，如：字际关系未经整理、字源未彻底分离、新旧字形并存、收字过滥等等，这些问题在一定程度上影响到字符集的科学性、系统性与利用价值。本文主要就字符集扩展 B 区（以下简称“扩 B”）所收汉字存在的具体问题提出一些粗浅看法，不当之处，敬请专家批评指正。

## 一、字形失真

汉字笔形有限、构成汉字的笔画数有限、构字部件有限、造字方法有限，在有限的基础上，字符集收字越多，字与字之间的区别率便会越低，区别率降低导致汉字错认、错写与错造。字符集基本集（以下简称“基本集”）中已有造错的字，如：基本集有“澁（6E8A）”字，此字未见传世字书，当是“澁”字误造（利用拼音输入法输入 bo 便会检索出“澁”字）。《汉语大字典[以下简称《大字典》]·水部》据《正字通》收“澁”字，同“波”，乃篆书“波”字楷定之异者（扩 B 收录“澁”[23D2B]，造成正误字形重出）。扩 B 收字多达四万二千七百多个，其中大部分已属俗讹字，这些“废字”是否有必要进入字符集，本文姑且不论，下面举几个扩 B 造错的字。

### 1、200C4

此字传世字书未见，当是“廐”字误造（字符集未收“廐”字）。《大字典·厂部》：“廐，同‘愈’。《字汇补·丿部》：‘廐，音义同愈。’”“廐”为“廐”字俗讹，《字汇补》、《大字典》谓“廐”同“愈”，非是（说详拙著《疑难字考释与研究》上编“廐”字条，P17）。

### 2、293B7

《篇海》卷十四《韋部》引《川篇》：“鞞，音亦，皮也。”《新修玉篇》卷二十六《韋部》引《类篇》字作“鞞”，音义同“鞞”。《大字典》据《篇海》收“鞞”字（《康熙字典·韋部补遗》同。《字海》收之，右上从夕），张涌泉先生谓“鞞”字当是“鞞”字俗书（《汉语俗字丛考》，P1132），所言是也（“鞞”字见今本《玉篇》，有音无义，恐亦俗讹字）。“鞞”字右旁乃“亦”字草书楷化之异者，扩 B 有“鞞”无“鞞”，“鞞”当是“鞞”字误为回改。

### 3、𩇛 29886

《大字典·馬部》五画内据《龙龕》收录“𩇛”字（《龙龕》卷二《馬部》字同），乃“𩇛”字俗讹。扩B有“𩇛”无“𩇛”，“𩇛”盖“𩇛”字误造。又：“𩇛”字五画，扩B列“駮（29885）”“馭（29887）”二字（皆四画）之间，归位亦有不妥。

### 4、駮 29909

《篇海》卷七《馬部》七画内引《搜真玉镜》：“駮，音驂。”（《篇海》同部七画内引《川篇》又收“駮”字，音桃，乃“駮”字俗讹）《大字典》据《篇海》收入而改从旧字形作“駮”，收于八画内。扩B收所“駮”字盖“駮”字误造。

### 5、駟 2989B

此字《大字典》据《说文》收录，《说文》十篇上《馬部》其字训马疾走，从馬，句声，当楷定作“駟”。字或作“駟”，“駟”字右旁所从与“駟”字右下所从相同。扩B有“駟”无“駟”，“駟”当是“駟”字之误，承《大字典》、《康熙字典》等传世字书而误。

### 6、𩇛 299CC

此字《大字典》据《篇海》、《字汇补》收录。《篇海》卷七《馬部》十画内引《搜真玉镜》字作“𩇛”，音讠，缺义训，其字下从木，不从术。“𩇛”字十四画，《篇海》列十画内，字形亦有误。扩B收“𩇛”字，亦承《大字典》而误。

以上六例，前四例属于扩B造字失真，后二例属于承前出字书而误。另外，

𩇛 𩇛 𩇛 𩇛 𩇛 𩇛  
扩B收 23AFD 23AFF 23B00 24CE1 26523 265F1 等字，不收 𩇛（小母鸡，音向，广

州方言。见《字海》P1009。𩇛与 𩇛 构字方法相同，未成人曰 𩇛，字同

“崽”、“仔”；未成母鸡曰 𩇛，俗体会意字也。）字，不知上揭诸字是何来源，

与“𩇛”字是否有关。

## 二、简化不当

扩B收录了部分类推简化字。学术界对于类推简化存有争议，本文姑且不论。

下面举出的是扩 B 所收因类推简化而造成的不当字形。

1、鬘 29B31 鬘 29B6F

“鬘鬘”之“鬘”已有简化字“丫”，无需类推出“鬘”字。

2、𩇛 299FB 𩇛 298F5

汉代西域大宛国出良马，后世因称良马为宛马；及“宛”用指马名，后世因造“𩇛”字。“𩇛”即“宛”之俗体加旁字。“𩇛”字如果有简化必要，亦应简化作马旁宛，扩 B 收“𩇛”字（右旁为“宛”之俗书），字形不当。

3、鬣 29BD2 鬣 29BF3 鬣 4C16

“鬣”为“鬣”字俗书，扩 B 收“鬣”之类推简化字，不收“鬣”之类推简化字。

4、𩇛 299F8 𩇛 298EB

扩 B 收“𩇛”字，而字符集中未见与之对应的繁体。扩 B 收“𩇛”字，不知与“𩇛”字是否有关系。

5、𩇛 293FE 𩇛 293A2

“𩇛”字从韋，支声，支字五画；扩 B 收“𩇛”字，左旁类推简化，右旁误作六画。

此外，扩 B 收“賈(27E52)”“賈(27E54)”一类简化字，不知是何来历。如果这些简化字来源于其他国家或地区，当与大陆汉字相分离，以避免干扰大陆汉字系统，避免汉字错误使用。

### 三、计画不确

字符集依部首排字，同部之字则依笔画多少排序，同笔画之字排在一起。扩 B 之中，多有计画不确而排序失次者，排序失次影响检索和利用。

1、𩇛 29903 𩇛 29937

扩 B “𩇛”字在八画内，“𩇛”在九画内。“𩇛”实有七画，比“𩇛”字少一画，反居“𩇛”字之后。

2、𩇛 2991C

此字《大字典》作“駮”，从“者”字旧字形，九画（《字海》改从新字形），扩 B 收“駮”字，实有八画，而列其字于九画，欠妥。

3、駮 2989D

此字四画，扩 B 列于“駮（2989C）”“駮（2989E）”二字之间，“駮”“駮”二字皆五画。

4、騫 2993A

《篇海》卷七《馬部》九画内引《搜真玉镜》：“騫，音余。”（“騫”为“騫”字俗讹）《大字典》据《篇海》收“騫”字，九画。扩 B 改作“騫”，八画，而仍收其字于九画，欠妥。

5、鬻 29BDD

此字《大字典》十三画，草头从旧字形作四画；扩 B 此字草头改作三画，实际笔画是十二画，仍收其字于十三画内。

6、鬻 29B91

此字《大字典》作“鬻”，九画；扩 B 改作“鬻”，实有八画，仍收其字于九画内。

#### 四、归部不当

归部不当亦影响检索和利用，扩 B 此类情况亦属常见。略举数例于下：

芒

20147

（见宀部。此为“芒”字异写，可不收；若收，当归艸部）

葱

22681

（心

甕

部。此为“葱”字俗书，当收艸部） 20085 （、部。此为“甕”字俗讹，当收宀

舟

部） 2053E

（冂部。同“终”，当收冂部）

趙

20804

（刀部。同“擣”，刺也。当收走

蹙

部） 214A5

（土部。“蹙”字俗讹，“蹙”或作“蹙”，同“蹙”。基本集“蹙”字



2、 𠂇 𠂇 𠂇 兵 𠂇 𠂇 𠂇 𠂇  
20505 20506 2050A 20503 20519 20517

3、 𠂇 𠂇 𠂇 𠂇  
20914 2091B 20924

4、 匚 匚 匚 匚 匚 匚  
20933 20934 20943 20977

5、 女 女 女 女 女 女 女 女  
216C0 216F7 216BF 216AB 217EE 21753 21730 2171F

6、 子 子 子 子  
21941 21967 21965 21975

7、 山 山 山 山  
21F40 21F11 21F76 21F0F

8、 戶 戶 戶 戶 戶 戶  
22A3E 22A47 22A44 22A33 22A42

9、 方 方 方 方 方  
230D8 230ED 230E5 23103

10、 𠂇 𠂇  
23137 23134

11、 木 木 木 木  
2338D 233D7

12、 𠂇 𠂇  
2388D 《《字符集》所收从毛之字，毛居左旁者均作半包围结构，如 23B23；此字作左右结构。《大字典》据《集韵》收录。宋刻本、扬州使院重刻本、文渊

阁四库全书本《集韵》，《五音集韵》、《篇海》引《馀文》并作“𤑔”，) 23C14 23C0F

13、 𤑔<sub>2453B</sub> 𤑕<sub>2453E</sub> 𤑖<sub>2453D</sub> 𤑗<sub>24529</sub> 𤑘<sub>2451C</sub>

14、 𤑙<sub>24579</sub> 𤑚<sub>2457B</sub>

15、 𤑛<sub>245FC</sub> 𤑜<sub>245EE</sub>

16、 𤑝<sub>24AF0</sub> 𤑞<sub>24AFA</sub> 𤑟<sub>24AEC</sub>

17、 𤑡<sub>24B31</sub> 𤑢<sub>24B28</sub> 𤑣<sub>24B2B</sub> 𤑤<sub>24BAE</sub> 𤑥<sub>24B6C</sub>

18、 𤑧<sub>24E4E</sub> 𤑨<sub>24DEB</sub> 𤑩<sub>24DEC</sub>

19、 𤑪<sub>24FE6</sub> 𤑫<sub>24FD7</sub>

20、 𤑬<sub>25940</sub> 𤑭<sub>259E1</sub> 𤑮<sub>2595B</sub> 𤑯<sub>259DA</sub> 𤑰<sub>259F3</sub>


21、 𤑲<sub>25FE0</sub> 𤑳<sub>260E1</sub> 𤑴<sub>26067</sub> 𤑵<sub>26038</sub> 𤑶<sub>260C3</sub> 𤑷<sub>26000</sub>

22、 𤑹<sub>26224</sub> 𤑺<sub>26250</sub> 𤑻<sub>26251</sub>







一画字中，实际十画。《大字典》二字收在十一画，皆从  𠄎

41、 𠄎 284FC 𠄎 284A0

42、 𠄎 25614 𠄎 25607 𠄎 25610 𠄎 25631

𠄎 25635 𠄎 25651 𠄎 256F4 𠄎 2564D

𠄎 25626 𠄎 2563D 𠄎 2567C 𠄎 2572C 𠄎 256D6

𠄎 200DE

43、 𠄎 2907F 𠄎 29078 𠄎 290C5

44、 𠄎 291D7 𠄎 291E4

45、 𠄎 291EB 𠄎 291F4

46、 𠄎 2968D 𠄎 296CF

47、 𠄎 29A80 𠄎 29A88 𠄎 29A72 𠄎 29A6B 𠄎 29A90 𠄎 29A6E 𠄎

48、 𠄎 29B04 𠄎 29B02 𠄎 29B8A 𠄎 29B2C

49、 𠄎 29C2E 𠄎 29C3D

𧈧 𧈨 𧈩 𧈪 𧈫 𧈬 𧈭

50、 29CF3 29CF4 29D1A 29C86 29C87 29D2B 29D30

𧈮 𧈯 𧈰 𧈱

51、 2A28B 2A293 2A2B2 2A2DF

𧈲 𧈳 𧈴 𧈵 𧈶 𧈷 𧈸 𧈹

52、 2A2FF 2A355 2A390 2A31B 2A38D 2A38E 2A35D 2A38F

𧈺 𧈻

53、 2A3B1 2A3B0

𧈼 𧈽

54、 2A3B5 2A3B7

比较扩 B 与基本集所收字形，不难发现基本集异形部首要比扩 B 少得多，有些在扩 B 存在异形的部首，在基本集中部首只有一种写法，如辵部扩 B 有七种写法，基本集只作辵；身部扩 B 有两种写法，基本集只作身；舟部扩 B 有五种写法，基本集只作舟；羽部扩 B 有两种写法，基本集只作羽等等。扩 B 部首异形有些是大陆不同历史时期汉字传承过程中形成的，也有些是因其他国家或地区与大陆汉字书写习惯不同造成的。无论哪种情况，都不仅仅是扩 B 所收的字存在部首异形，基本集所收的字在不同历史时期（或不同字书）、不同国家或地区之间部首也存在异形现象。如：基本集所收从辵之字，《康熙字典》皆从辵，台湾地区皆从辵；所收从骨之字，《康熙字典》皆从骨，台湾地区皆从骨。基本集收录的字，使用频率要远远大于扩 B 收录的字（基本集收录的字有的一天使用数以亿计，扩 B 收录的字有的若干年可能连一次也用不到），既然基本集不照顾部首时空差异，扩 B 照顾部首的时空差异又有什么意义？

### 七、同画无序

字符集依据部首笔画排序，同部同笔画的字则排列无序。字符集不断扩展，至扩 B 收字已近七万，其中有的部首同部同笔画字多达数十个、上百个，检索殊为不易（下表为扩 B 彡部八画字排列情况）。建议参照大型字书编排方法，同部同笔画字依起笔笔顺横、竖、撇、点、折排序，同笔画且起笔笔顺相同者，依第二笔笔顺排序。

鬃 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚  
2985E 2985F 29860 29861 29862 29863 29864 29865 29866 29867 29868 29869 2986A

鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚 鬚  
2986B 2986C 2986D 2986E 2986F 29870 29871 29872 29873 29874 29875 29876 29877 29878

鬚  
29879