## Glyph corrections for U+027F and U+0285 in TUS

Lorna A Priest and Greg Aumann

SIL International
2007-Jan-12

## Introduction

The charts for The Unicode Standard show U+027F LATIN SMALL LETTER REVERSED R WITH FISHHOOK and U+0285 LATIN SMALL LETTER SQUAT REVERSED ESH with the glyphs on the baseline (see Figure 1). Chinese books consistently use these characters with the glyph descending below the baseline as shown below and in Figures 2-7. Figures 2b and 4b are a magnification of one portion of Figures 2a and 4 a respectively to show clearly where the baseline is in relation to these characters. We would like to request the UTC to correct the glyphs in the charts and elsewhere where they may be referenced.

## baseline 1 〕 baseline

The symbols were invented by Karlgren about 100 years ago. Figure $2 b$ shows how he used the characters with the glyph descending below the baseline.

Both characters were added to Unicode 1.1. The shape of U+0285 LATIN SMALL LETTER SQUAT REVERSED ESH was changed between version 2.0 and 3.0 of the Unicode standard to be "squat." Since the name has not changed it's probable the original samples for U+0285 were "squat" and the glyph in TUS was just wrong from the beginning.

We have been unable to find the document requesting the glyph change for $\mathrm{U}+0285$ or the original document proposing these characters. If the original proposal used older printed materials the baselines do jump around and if there are no characters immediately adjacent it would be very difficult to tell where the baseline is intended to be. The examples in this document are words with lots of adjacent characters and recent versions so it is much clearer where the baseline is.

Figures


Figure 1. TUS. p. 591.
A. royelle apico-gingivale, haute, tendue, délabialisée ou à
Touverture labiale large, se rencontre en Mand. Yangtc. Wou Yue, peut-être Swat. ${ }^{1}$ Jap. Elle est partout orale, et n'apparaît qu'en syllabe ouverte et après s, z, p. ex. Pék. sl, hch. sseu 'mourir'. Edkins compare cette voyelle au 'e' angl. de tassel, comparaison assez heureuse en considérant que dans ce mot la langue, à cause du 1 suivant, ne s'éloigne que très peu de la position du s. L'absence d 1 ans les autres langues connues rend sa définition très difficile pour des amateurs. Les identifications courantes comme \&l'u bref anglais», «l'eu français etc. sont toutes très incorrectes.
, voyelle apico-alvéolaire, haute, tandue, délabialisée ou à l'ouverture labiale large existe en Pék. Chansi (sauf T'aiy. T'aik. et P'ingy.) Kansou Chensi Hoaik. K'aif. Sseu. Nank. Elle est partout orale et n'existe qu'en syllabe ouverte après $\mathbb{s}, \boldsymbol{z}$ p. ex. Pék sl, H.ch. che 'temps'. C'est probablement aussi cette voyelle qu'a eu en vue M. Parker, quand il écrivait chï, chïng pour le dial. de Ningpo. Ce qui est dit ci-dessus des définitions de la voyell 1 sapplique aussi bien à celles de $\imath$, publiées jusqu'ici. Une voyelle très analogue \&e rencontre dans certains dial. suédois; on l'appelle \&i de Vibys.
Les voyelles 1 et $l$. gnt cause de l'embarras aux transcripteurs. Les russes ond choisi la lettre ы pour toutes les deux, graphie assez bonne au point de vue acoustique. Le parti pris par M. Mateer, qui écrit '"' pour toutes les deux est également bon. La
${ }^{1}$ Cf. ba de la série vélo-palatale (p. 308).
Figure 2a. Karlgren. p. 295.
Les voyelles 1 et 2 ont cansé de l'embarras aux transcripteurs. Les russes ont choisi la lettre m pour toutes les deux, graphie assez bonne au point de vue acoustique. Le parti pris par M. Mateer, qui écrit '..’ pour toutes les deux est également bon. La

Figure 2b. Karlgren. p. 295 (showing baseline).

Table 2. Finals

> 2(a) In G. R.


## 2(b) In IPA

| C | Open Ending | -i | -u | -n |  | - $\eta$ |  | $-\lambda$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 123 | 45 | $6 \quad 7$ | $8 \quad 9$ | 10 | 11 | 12 | 13 |
| Row-a | $\underset{1}{z,} \mathrm{x}_{1}^{*}=$ A | ai ei | au ou | an on |  | $\wedge \eta$ | - $\eta$ | $\boldsymbol{\gamma}$ |
| Row-i | i iA i $\varepsilon$ | iai | iau iou | ien in | iaŋ | i | iదך |  |
| Row-u | u uA $\mathrm{ur}^{\text {¢ }}$ | uai uei |  | uan uən | uaj | $u \wedge \eta$ |  |  |
| Row-iu | $\mathrm{y} \quad \mathrm{y} \varepsilon$ |  |  | yan yn |  |  |  |  |

${ }^{\circ}$ Karlgren invented the symbols 1 and $\eta$ for these apical vowels, as he calls them. See his Phonologie, pp. 295-297, Chinese transl., pp. 197-199.

Figure 3. Chao. 1968. p. 24.
29.

30．d
31．$s$
32． z
33．c
34． c
35．J
36． n
37．
38． j
$j$ 广州：夜［ $\mathrm{j} \varepsilon-1$ ］
39．t6 北京：鸡［tci7］
40．t6 ${ }^{6}$ 北京：欺［ $\mathrm{tc}{ }^{\text {＇}} \mathrm{i} \mathrm{C}$ ］
41． dz 苏州：骑［dzi4］
42．no 苏州：银［nin1］

43． 6 北京：希［6i7］
44．z 浙江绍兴：自［zi」］
45． k 北京：该［kai］
46． $\mathbf{k}^{\star}$ 北京：开［k＇ai］
47． g 苏州：茄［gD1］
48．$刀$ 苏州：误［ŋouv］
49． x 北京：海［xai＾］
50． X 湖南双峰：鞋［ $\mathrm{ya-1}$ ］
51．$x$ 山西太谷：胡［ $x u-1$ ］
52．？云南玉溪：肝［？ $\mathfrak{\sim}-1$ ］
53．$P$ 云南玉溪：宽［ ${ }^{\prime}$ ‘ũ－1］
54． h 苏州：花［ho－1］
55．Һ
56．清音化：b，d。

四，元音举例

1． 1 北京：丝［ s 17$]$
2． 2 北京：诗［sนา］
3． 4 苏州：书 $[s q-1]$
4． 4 湖北麻城：树［\＄Ч－1］
5．i 北京：衣［i7］
6． 1 苏州：烟［i1－1］
7．e 北京：梅［mei1］
8．E 苏州：三［sE－1］
9．$\varepsilon$ 广州：爹［tะา］
10．$æ$ 苏州：好［hæu］
11．a 北京：干［kan7］，广州：花［faา］
12．A 北京：家［tcia7］
13．B 广州：闭［pei－1］
14．a 北京：纲［kaŋา］
15．D 苏州：街［kD－1］，南京：茶［ts＇D1］
16.

17．$o$
． 6 ．

18．$u$
19． u
20．y
21．
22．œ
23． y
24． u
25．ar
26．$\Theta$［広州：虚［heyา］
27．～鼻化，厦门：影［さ̃̃v］
28．or
29．or 估较低
30．os 唇较圆
31．os
32． 0
33．or
34．or

舌较高
湖南双峰：波［pu7］
北京：乌［u7］
北京：雨［yA］
苏州：干［kø1］
广州：靴［hœา］
北京：饿［8v］
兰州：二［ $\left.\mathrm{\gamma u}_{1}\right]$
北京：儿［ $\boldsymbol{\sigma}_{1}$ ］
广州：虚［hey7］

唇较展
开口度较小
舌位较前
舌位较后

其 t Trit
F taty

Figure 4a．Yuan．1989．p． 6.


Figure 4b. Yuan. 1989. p. 6 (magnification showing baseline).

## 8．两种来源的 $\mathrm{pf}-, \mathrm{pf}^{4}-$

pf－，pff－声母出现在西北。按来源可以分两种情况。
（1）帮组来源的 $\mathrm{pf}-, \mathrm{pf}^{4}-$ ——陕西宝鸡，麟游等地， u 韵和 o 韵之前没有 $\mathrm{p}-, \mathrm{p}^{4}$－，只有 $\mathrm{pf}-$ ， pf＂－。如＂跛＂pfo，＂坡＂pf＇o，＂布＂pfu，＂朴＂${ }^{2} f^{4} u_{0}$（
（2）知照组来源的 $\mathrm{pf}^{2}, \mathrm{pf}^{4}-—$ 凡北京 $\mathrm{ts}-, \mathrm{ts}$＇－声母的 $u$ 韵字，在兰州，西安，潼关等地都是 $\mathrm{pf}-, \mathrm{pf}^{\prime}$－声母的 u 韵字；凡北京 $\mathrm{ts}-$ ， $\mathrm{ts}^{\mathrm{s}}$－声母的 u 韵以外的合口字，在上述地区都是 $\mathrm{pf}^{\mathrm{f}}$ ， $\mathrm{pf}^{\mathrm{d}}$－声母

 ＂软＂ v 笈。但知照组来源的开口字，在 $\downarrow$ 和 $\gamma$ 二韵前读 $t s-t s^{4}$ ，在其他韵母前读 $t-t^{4}$ ，例如
 $\mathrm{t}^{\mathbf{t}}$－也已变成 $\mathrm{ts}-\mathrm{ts}{ }^{\mathbf{s}}$ ，二者的区别渐趋泯灭。 ${ }^{(2)}$

山西永济，运城也有 $\mathrm{pf}^{2}$ ， $\mathrm{pff}^{4}$－声母，（3）情况与西安相似。
9．＂资雌私＂和＂知虫诗＂声母的分混
这两组字的声母在北京话是有分别的，＂资雌私＂是 $\mathrm{ts}-, \mathrm{ts}{ }^{4}, \mathrm{~s}$－，＂知革诗＂是 $\mathrm{ts-}, \mathrm{ts}{ }^{\mathbf{s}} \mathrm{-}, \mathrm{~s}-$ 。多数地区这两组声母有分别，只是分别的具体情况，即哪些字属 ts －组，哪些字属 ts －组，不全一样。完全不分的地区比较少些。大体上说，北方方言中，除东北以不分占优势（如沈阳，丹东，哈尔滨）外，一般是能分的。西北方言中，太原，沁县等地不分，其他很多地区能分。西南方言绝大部分地区不分，少数地区如昆明，楚雄能分。江淮方言一般不分，南京，巢县等地能分。

各种分混情形列表如下（以 s －和 s －的分混为例）：

|  | 送（心母） | 生（审二） | 诗（审三） | 身（审三） |
| :---: | :---: | :---: | :---: | :---: |
| 北京 | suy ${ }^{\text {？}}$ | （s）n | s？ | （son |
| 当阳（湖北） | şup ${ }^{2}$ | （son | cs？ | （\＄ŋn |
| 昆明 | son ${ }^{\text {a }}$ | cs® | cs？ |  |
| 天津 | sun | （s）n | cs1 | （son |
| 成都 | son ${ }^{2}$ | cson | cs1 | cson |

10．日母在各地的变化
日母字声母（这里不包括止摄开口三等字＂儿耳二＂等的读音）大部分地区一律是 $z_{-}$，或者一律是 $\mathrm{z}_{-}$。一般是有 ts －组的地区是 $\mathrm{z}_{-}$－如北京，否则是 $\mathrm{z}_{-}$。

有些地区情况不同。比如湖北当阳为 $\emptyset$－；扬州＂热＂是 $\emptyset$－，其他一律是 $1-$ ；沈阳为 $\emptyset$－（齐齿或撮口）；济南开口韵前为 $z_{-}$，合口韵前为 $1-$ ；汉口合口撮口韵前和＂日＂是 $\emptyset$－，开口韵前是 $n$－；西安，兰州等地合口韵前是 v－（参看第8节），开口韵前是 $\mathrm{z}_{\mathrm{H}}$（但西安＂日＂ $\boldsymbol{r}$ ）。如下表：


[^0]Figure 5．Yuan．1989．p． 31.

| $\begin{array}{\|cccc} \hline & & & \\ \text { 方 } & & \text { 相 } \\ & & \text { 中 } & \\ & \text { 音 } & & \text { 古 } \\ \text { 方言 } & \text { 点 } & & \\ \hline \end{array}$ |  | 肆放碳 | 似 | 祀 | 寺 | 饲 | 知知这 | 蜘 | 支 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $\begin{aligned} & \text { 止开三 } \\ & \text { 去至 } \end{aligned}$ | $\begin{aligned} & \text { 止开三 } \\ & \text { 上止相 } \end{aligned}$ | $\begin{aligned} & \text { 止开 } \equiv \\ & \text { 上止弱 } \end{aligned}$ | 止开 沫 <br> 去志雅 | $\begin{aligned} & \text { 止 开三 } \\ & \text { 志志邪 } \end{aligned}$ | $\begin{aligned} & \text { 止开三 } \\ & \text { 平支知 } \end{aligned}$ | 止开三 平支知 |  |
| 北 | 京 | s1 ${ }^{\circ}$ |  | S1 ${ }^{\circ}$ | si ${ }^{\circ}$ | S1 ${ }^{\circ}$ | ．tst | ．tst | ．tst |
| 济 | 南 | s1 ${ }^{\text {a }}$ |  | s1 ${ }^{\prime}$ | s1 ${ }^{\text {a }}$ | s1 ${ }^{1}$ | tst | ts？ | ． t \＄1 |
| 西 | 安 | s1 ${ }^{\circ}$ | si si \％ | S1 ${ }^{1}$ | s1 ${ }^{\circ}$ | S1 | ．tst | tst | ts ${ }_{1}$ |
| 太 | 原 | s1 ${ }^{\circ}$ | s1 ${ }^{\circ}$ | s1 ${ }^{\circ}$ | s1 ${ }^{\circ}$ | ${ }^{\text {S }} 1$ | ．ts 1 | ．ts1 | ．ts1 |
| 武 | 汉 | s1 ${ }^{\circ}$ | s1． | s1 ${ }^{\circ}$ | si ${ }^{\text {a }}$ | $\begin{aligned} & \mathbf{s i}^{\prime} \\ & \text { ts }^{\prime}{ }_{11} \end{aligned}$ | .$_{\text {ts }}^{1}$ | .$_{\text {ts }}$ | ．ts ${ }_{1}$ |
| 成 | 都 | s1 ${ }^{\circ}$ | s1 ${ }^{\prime}$ | S1 ${ }^{\circ}$ | s1 ${ }^{\circ}$ | S1 ${ }^{\circ}$ | ts ${ }_{1}$ | ．ts1 tse谷 | ．ts 1 |
| 合 | 肥 | s1 ${ }^{1}$ | s1 ${ }^{2}$ | s $1^{\circ}$ | s1 ${ }^{\circ}$ | s1 ${ }^{\prime}$ | ．tst | $\begin{aligned} & \text { tsl } \\ & . \operatorname{tse} \text { 俗 } \end{aligned}$ | ．tst |
| 扬 | 州 | si | s1 ${ }^{\circ}$ | si ${ }^{\circ}$ | si ${ }^{\circ}$ | s1＇ | ．ts1 | ts ${ }_{1}$ | ．ts1 |
| 苏 | 州 | si ${ }^{\prime}$ | $z^{1}$ | $z_{1}{ }^{2}$ | $z_{1}{ }^{2}$ | $z_{1}{ }^{\circ}$ | ，tsy | tsy | ．tsy |
| 温 | 州 | si ${ }^{\circ}$ | ＇z1 | ＇z1 | $z_{1}{ }^{2}$ | $z_{1}{ }^{\circ}$ | ．ts1 | ．ts ${ }_{1}$ | ．tsei |
| 长 | 沙 | s1 ${ }^{\circ}$ | $\begin{aligned} & \mathrm{s} 1^{\circ} \text { 文 } \\ & \mathrm{s} \mathbf{n}^{\prime} \end{aligned}$ |  | $\mathrm{ts}_{1}{ }^{2}$ | s1 ${ }^{\circ}$ | ，ts1 | $\begin{aligned} & \mathrm{ts1} \\ & \mathrm{ts} 1 \\ & \mathrm{ts}, \end{aligned}$ | ．tsl |
| 双 | 峰 | s1 ${ }^{\circ}$ | $\mathrm{dz} 1{ }^{2}$ | $\mathrm{dz}_{1}{ }^{\circ}$ | $\mathrm{dz}_{1}{ }^{\prime}$ | $\mathrm{dz}_{1}$ | tst | tst | ．tst |
| 南 | 昌 | s1＇ | si ${ }^{\circ}$ | $\mathrm{sic}^{2}$ | s1＇ | S1 ${ }^{\text {P }}$ | ts 1 | tst | ．ts1 |
| 梅 | 县 | si | si ${ }^{\circ}$ | $\mathrm{si}^{\prime}$ | ts＇ $1^{\prime}$ | ． 81 |  | ${ }_{\text {ts }} 1$ | .$^{\text {ts }} 1$ |
| 广 | 州 | fi | ＇ts＇i | tji | $t \mathrm{i}$ \％ | t5i | t $\mathrm{t} \mathbf{i}$ | ．tfi | ．t5i |
| 阳 | 江 | \＄ei ${ }^{\text {c }}$ | $\begin{aligned} & t \int \mathrm{ei}^{\prime} ⿱ 亠 乂 \\ & { }^{t} \int^{\prime} \mathrm{ei} i_{1} \\ & \hline \end{aligned}$ | $t$ ¢ $\mathrm{i}^{2}$ | $t$ ¢ $\mathrm{i}^{\text {P }}$ | $t_{\text {fei }}{ }^{2}$ | t t i | ．t5i | ．tsi |
| 庢 | 门 | su＇ | $\text { su' }{ }^{\text {x }}$ $\mathrm{sai}^{\prime} \mathrm{g}$ | $s^{*}{ }^{*}$ x <br> sai ${ }^{\text {自 }}$ | si ${ }^{\text {P }}$ | $\mathrm{su}^{+}$文 <br> ts $\mathrm{i}^{\prime}{ }^{\prime}{ }^{\text {自 }}$ | $\begin{aligned} & \mathrm{tin}^{\left(\mathrm{t}^{(3)}\right.} \\ & \mathrm{ti}^{\frac{1}{x}} \end{aligned}$ | ．ti | ．tsi 文 <br> ，kí白 |
| 潮 | 州 | si ${ }^{\circ}$ | ＇s1 | ${ }^{\text {＇s }} 1$ | $\mathrm{zi}^{\text {e }}$ |  | ti＇文 tsai | ．ti | ．tsi |
| 福 | 州 | $s ø y^{\prime}$ | søy ${ }^{2}$ | $\begin{aligned} & \begin{array}{l} \text { soy }{ }^{\prime} \text { x } \\ \text { sai' } \end{array} \end{aligned}$ | seit | søy＇＊ <br> ts＇ei 白 | ${ }_{\text {，} \mathrm{ti}} \mathrm{l}$ | ．t＇i | ．tsie |
| 建 | 怄 | si ${ }^{\text {a }}$ | su＇ | su＊ | tsi ${ }^{\text {e }}$ | $\mathrm{si}^{+}$ | ．ti | ．ti | ．tsi |

（1）$s i^{\circ}$ 为＂是＂训读。（2）又通＂智＂，＊知义切，止开三去寘知。（3）tsai白。

Figure 6．Linguistics Section．1989．p． 59.

Table 8.5. Development of the Middle Chinese final consonants in Mandarin dialects

|  | Middle <br> Chinese | Peking | Yángzhōu | Xīā | Chéngdū | Língbăo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 'three' | sâm | $\operatorname{san}^{1}$ | sã ${ }^{1}$ | sã ${ }^{1}$ | $\operatorname{san}^{1}$ | sã ${ }^{1}$ |
| 'umbrella' | sân: | $\operatorname{san}^{3}$ | s $\tilde{x}^{3}$ | sã ${ }^{3}$ | $\operatorname{san}^{3}$ | sã ${ }^{3}$ |
| 'mulberry' | sâng | san ${ }^{1}$ | say ${ }^{1}$ | $\operatorname{san}^{1}$ | $\operatorname{san}^{1}$ | san ${ }^{1}$ |
| 'twelfth month' | lâp | $1 \mathrm{a}^{5}$ | næP ${ }^{7}$ | $1 \mathrm{la}^{1}$ | na ${ }^{2}$ | $1 \mathrm{la}^{1}$ |
| 'pungent' | lât | $1 a^{5}$ | næ? ${ }^{7}$ | $1 \mathrm{l}^{1}$ | na ${ }^{2}$ | $1 \mathrm{l}^{1}$ |
| 'drop' | lâk | $140^{5}$ | nap ${ }^{7}$ | luo ${ }^{1}$ | $\mathrm{no}{ }^{2}$ | $1 u^{1}$ |

Leaving aside the medials and those finals which contain a single high vowel, most Mandarin dialects show a basic vocalic contrast between two degrees of openness, usually represented as $a$ and $\partial$; while other non-high vowels sometimes occur, they do not ordinarily contrast with $a$ and a phonemically. We will see soon that a three-way contrast of non-high vowels is one of the features of the Central dialects.

Apical and retroflex vowels are a feature found in many Chinese dialects, but are especially prevalent in the Northern and Central groups. In addition to the unrounded apica1 [1] and the retroflex [ [7] vowels described for the standard language, a few dialeets also possess rounded counterparts, a rounded apical vowel $[\Psi]$ and a rounded retroflex vowel $[\Psi]$ respectively.

Dialects of the Northern zone are generally rich in diphthongs and triphthongs, in conspicuous contrast to the Wú dialects, for example, where such combinations are much rarer. This may be viewed as a conservative feature when compared with Middle Chinese, in which vocalic clusters abounded.
From a comparative point of view, Mandarin dialects have a small number of tones. A few dialects have only three tones, such as Línchéng in Héběi. More typically four tones are found, as in the Peking dialect. In the Jiāng-Huái region five tones are not uncommon. A few Mandarin dialects with six tones have been reported along the boundary that separates the Northern and the Central dialects. Many Mandarin dialects (perhaps a majority) show a tonal pattern similar to that described for Peking: the ping tone occurs in two registers, the yinping in words which in Middle Chinese had voiceless initials, and the yángping in words which in Middle Chinese had voiced initials: here and there one can encounter dialects in which these two categories have merged, for example, Zhāngiiākǒu in Héběi and Tàiyuán in Shānxī. No Mandarin dialect (as far as is known) shows a register distinction for the shăng tone; words with voiced obstruent initials of this

Figure 7. Norman, Jerry. 1988. p. 194.

## References

Chao，Yuen Ren．1968．A Grammar of Spoken Chinese．Berkeley and Los Angeles： University of California Press．
Karlgren，Bernhard．1915－1926．Etudes sur la Phonologie Chinoise．Archives D＇etudes Orientales，in 4 volumes，Vol．15：1－4．Published by J．－A．Lundell．Brill， Leiden／Uppsala．
Linguistics Section，Department of Chinese Language and Literature，Beijing University （北京大学中国语言文学系语言学教研室）（ed．）．1989．Hanyu Fangyin Zihui （汉语方音字汇）。［Phonetic Dictionary of Chinese Dialects］．Second edition． Beijing：Wenzi Gaige Publishing House［文字改革出版社］．（The first edition was published in 1962．）
Norman，Jerry．1988．Chinese．Cambridge Language Surveys series．Cambridge，UK： Cambridge University Press．
The Unicode Consortium．2007．The Unicode Standard，Version 5．0．0，defined by：The Unicode Standard，Version 5．0．Boston，MA：Addison－Wesley．
Yuan，Jiahua（袁家骅）et．al．1989．Hanyu Fangyan Gaiyao（汉语方言概要）［Outline of Chinese Dialects］．Second edition．Beijing：Yuwen Publishing House［语文出版社］．（The first edition was published in 1960．）


[^0]:    （1）参见白涤洲，喻世长（关中方音调查报告）第 6 页，中国科学院出版， 1954 年。
    （2）同上。
    （3）根据高本汉（中国音韵学研究）第 174 页，商务印书馆， 1940 年。

