Proposal to encode Javanese and Sundanese Arabic characters

Denis Moyogo Jacquerye <moyogo@gmail.com>  
2019-10-03

1. Introduction

The Arabic alphabet has been adapted and used to write languages in the Malay Archipelago since at least the 14th century. The most used adaptation is the Jawi alphabet, used in Classical Malay and several Modern Malayic languages such as Malaysian Malay (zms), Indonesian Malay (id), Brunei Malay (kxd), Acehnese (ace), Banjar (bjn), Bengkulu (pse), Minangkabau (min), Musi (mui) and others.

Other languages in the region also use adaptations of the Arabic alphabet, sharing letters with Jawi but also using variants or additional letters and diacritics:

- Pego is the Arabic alphabet used for Javanese (jv), Sundanese (su) and Madurese (mad), with some variation between those languages or between works,
- Buri Wolio is the Arabic alphabet used in Wolio (wlo),
- Sulu or Sulat Sug is the one used for the Tausug language (tsg),
- Serang is the one used for Makassarese (mak).
Figure 1: Map of languages of the Malay Archipelago using the Arabic alphabet in Cho (2018), showing Jawi, Pegon, Serang, Wolio and Sulat Sug (Tausug).

Javanese Pegon uses \texttt{ARABIC LETTER TAH WITH THREE DOTS BELOW} or \texttt{ARABIC LETTER TAH WITH DOT BELOW}, which are proposed to be encoded. Malagasy (mg) Sorabe, the Arabic alphabet used in Malagay in Madagascar, also uses \texttt{ARABIC LETTER TAH WITH DOT BELOW}.

Sundanese Pegon uses the \texttt{ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW} which is proposed to be encoded.

In Gallop et al. (2015) p. 14, it is noted that “the Malay ga usually has one or three dots above or below”. Werndly (1736), p. 6, describes such reduction of three dots noting it doesn’t occur with letters that would clash with letters in the same shape group that have a single dot, such as \texttt{0686 ARABIC LETTER TCHEH چ} and \texttt{062C ARABIC LETTER JEEM ج}, \texttt{06A4 ARABIC LETTER VEH ف} and \texttt{0641 ARABIC LETTER FEH ف}, or \texttt{06A0 ARABIC LETTER AIN WITH THREE DOTS ABOVE ڠ} and \texttt{063A ARABIC LETTER GHAIN غ}. This variation of number of dots and their position in Jawi manuscripts can also be observed in some Javanese or Sundanese Pegon works, for \texttt{ga (KAF- or KEHE-shape)}, \texttt{dha (DAL-shape)} or \texttt{tha (TAH-shape)}, with three dots below or one dot below.
2. Proposed characters to be encoded

ARABIC LETTER TAH WITH DOT BELOW and ARABIC LETTER TAH WITH THREE DOTS BELOW

Many Javanese Pegon works use variants of a dotted TAH to represent a voiceless retroflex stop /ʈ/: ARABIC LETTER TAH WITH DOT BELOW or ARABIC LETTER TAH WITH THREE DOTS BELOW, which need to be encoded. It also uses 068A ARABIC LETTER DAL WITH DOT BELOW or 08AE ARABIC LETTER DAL WITH THREE DOTS BELOW to represent a voiced retroflex stop /ɖ/. Adelaar (1995), p. 337, notes that ARABIC LETTER TAH WITH DOT BELOW and 068A ARABIC LETTER DAL WITH DOT BELOW are also used for the same sounds in Sorabe, the Malagasy (mg) Arabic alphabet, in Madagascar.

<table>
<thead>
<tr>
<th>codepoint</th>
<th>isolated</th>
<th>final</th>
<th>medial</th>
<th>initial</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+088B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARABIC LETTER TAH WITH DOT BELOW</td>
</tr>
<tr>
<td>U+088C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ARABIC LETTER TAH WITH THREE DOTS BELOW</td>
</tr>
</tbody>
</table>
ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW

Various Sundanese Pegon works use ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW, for /g/, which needs to be encoded.

<table>
<thead>
<tr>
<th>codepoint</th>
<th>isolated</th>
<th>final</th>
<th>medial</th>
<th>initial</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+088D</td>
<td>ⵜ ⵜ ⵜ ⵜ</td>
<td>ⵜ ⵜ ⵜ ⵜ</td>
<td>ⵜ ⵜ ⵜ ⵜ</td>
<td>ⵜ ⵜ ⵜ ⵜ</td>
<td>ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW</td>
</tr>
</tbody>
</table>

3. Unicode character properties

088B;ARABIC LETTER TAH WITH DOT BELOW;Lo;0;AL;;;;;N;;;;;
088C;ARABIC LETTER TAH WITH THREE DOTS BELOW;Lo;0;AL;;;;;N;;;;;
088D;ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW;Lo;0;AL;;;;;N;;;;;

4. Joining type and group for ArabicShaping.txt

088B; ARABIC LETTER TAH WITH DOT BELOW; D; TAH
088C; ARABIC LETTER TAH WITH 3 DOTS BELOW; D; TAH
088D; ARABIC LETTER KEHEH WITH VERTICAL 2 DOTS BELOW; D; GAF

5. Annotations

The following annotations could be updated in NameList.txt.

068A ARABIC LETTER DAL WITH DOT BELOW
* Pegon, Malagasy
08AE ARABIC LETTER DAL WITH THREE DOTS BELOW
* Belarusian
* Pegon alternative for 068A

The following annotations are recommended for NameList.txt.

0 Additions for Pegon:
088B ARABIC LETTER TAH WITH DOT BELOW
* Pegon, Malagasy
088C ARABIC LETTER TAH WITH THREE DOTS BELOW
* Pegon alternative for 088B
088D ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW
* Sundanese Pegon

6. Suggested collation

088B ARABIC LETTER TAH WITH DOT BELOW and 088C ARABIC LETTER TAH
WITH THREE DOTS BELOW can be sorted after 08A3 ARABIC LETTER TAH WITH TWO DOTS ABOVE and before 0639 ARABIC LETTER AIN.

088D ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW can be sorted after 063B ARABIC LETTER KEHEH WITH TWO DOTS ABOVE.
Berdasarkan padanan aksara Jawa dan Pegon, tampaklah bahwa aksara Pegon yang bentuknya huruf Arab mempunyai sistem bunyi yang sama dengan aksara Jawa. Namun, dari padanan aksara Arab dan Pegon, terlihat bahwa huruf Pegon yang bunyi dan bentuknya sama dengan huruf Arab dalam sistem tulisan Arab jumlahnya hanya tiga belas, yaitu ba (ب)، ta (ت)، jim (ج)، dal (د)، ra (ر)، sin (س)، kaf (ك)، lam (ل)، mim (م)، nun (ن)، wau (و)، alif (أ)، dan ya (ي)، sedangkan tujuh aksara Pegon lainnya, yaitu ca (چ)، pa (پ)، dha (ذ)، nya (ن)، ga (گ)، tha (ث)، dan nga (ڠ).

Ketujuh huruf Arab hasil modifikasi yang bentuknya mirip dengan huruf Arab, tetapi terdapat tanda titik diakritik dan bunyi-nya tidak dikenal dalam sistem tulisan huruf Arab, disebut huruf Arab rekaan. Ketujuh huruf Arab rekaan ini, kira-kira memang diciptakan untuk mewakili bunyi-bunyi yang ada dalam bahasa Jawa, karena tidak ada padanan bunyi-nya dalam sistem tulisan Arab. Bagaimana ketujuh huruf Arab rekaan ini diciptakan hingga kini belum diketahui dengan pasti. Namun, bertolak dari pendapat Achadiati Ikram (1977:38) yang mengatakan bahwa Jawi adalah aksara Arab yang digunakan sebagai sarana rekam bahasa Melayu, maka sistem tulisan Pegon dipadankan dengan Jawi. Berdasarkan padanan itu, dapat diketahui bahwa lima dari tujuh huruf Arab rekaan Pegon, yakni ca (چ)، pa (پ)، nya (ن)، ga (گ)، dan nga (ڠ) juga terdapat di dalam tulisan Jawi.

Figure 3 : Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Yulianto and Pudjiastuti (2001), p. 207.
The Pegan script has similarities and differences with Arabic script and Jawi script that developed and used in Malaysia, Brunei, Singapore and Pattani Thailand. The equation is: (1) The script is written from right to left. (2) In the script system, there are alphabets arranged with each other. (3) There is a dot sign and a harak or syakal. (4) There is a change of alphabets when in front, middle and back. (5) The alphabet connection is in one word, not between words. (6) In one word, there can be several pieces of alphabet connection. (7) Numbers are written from left to right. While the difference is: (1) System script in Arabic Pegan is based on the phonological system of Java language. (2) The vowels in the Pegan script are more than the vowels contained in the Arabic script as well as the Jawi script. (3) In the Pegan script, there is a syllable, while in Jawi writing does not exist. (4) There are alphabets غ،ط،ك،ي،د،ف،ج، which are not found in Arabic script. (5) There are alphabets د and ب، which are not found in Jawi. (6) The sound /g/ is written with the alphabets للاء، while in Jawi it is written with ِإ. (7) The sound /p/ written with the alphabet ض، while in Jawi script written with ض.
Figure 6: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Semarang (19??), p.12.

Figure 7: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Anwar (19??), p.8.

Figure 8: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Anwar (19??), p.28.
Figure 9: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Musthofa (19??), p.3.

Figure 10: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Asnawi Kudus (190?), p.8.

Figure 11: Javanese Pegon ARABIC LETTER TAH WITH DOT BELOW in Asnawi Kudus (190?), p.48.
It is also necessary to draw the adaptations of the Arabic script as used for some other Indonesian languages into one’s comparison. Pegon, for instance, shows remarkable agreement with Sorabe in that it uses the same graphic conventions for retroflex $d$ and $t$ as does Sorabe for $d$ and $t$: retroflex $d$ is written as a dal with a subscript dot (۰۰), and retroflex $t$ is written as a ta [ta] with a subscript dot (٠٠). For a better appreciation of the agreement it should be added that Javanese speakers often perceive the alveo-

Figure 12: Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW in Adelaar (1995), p.337.

Figure 13: Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as ۰۰۰۰ ARABIC LETTER DAL WITH DOT BELOW (not marked) in MS 18141, f.1r.
Figure 14: Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as 068A ARABIC LETTER DAL WITH DOT BELOW (not marked) in Manuscrit Malayo-polynésien 20 f.4.

Figure 15: Malagasy Sorabe ARABIC LETTER TAH WITH DOT BELOW as well as 068A ARABIC LETTER DAL WITH DOT BELOW (not marked) in Manuscrit Malayo-polynésien 26 f.3.
Figure 16: Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in MS 12305 (late 18th, early 19th century), f.4v.

Figure 17: Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Mustofa 1960, p.1.
Figure 18: Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), Syiir Sekar Cepogo, p.6.

Figure 19: Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), Syiir Sekar Cepogo, p.9.
Figure 20: Javanese Pegon ARABIC LETTER TAH WITH THREE DOTS BELOW in Al-Hajawi (19??), Syiir Sekar Kedathon, p.12.
**Figure 21**: Sundanese Pegon Arabic Letter Keheh with Two Dots Vertically Below in Coolsma (1985), p. 10.
Figure 22: Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in EAP211/1/3/2 p. 198.

Figure 23: Sundanese Pegon ARABIC LETTER KEHEH WITH TWO DOTS VERTICALLY BELOW in British and Foreign Bible Society (1930), The Gospel in many tongues, p.128.
8. Bibliography


Raden Muhammad Nuh Cianjur (19??). Risalah qiblat diterangkan dengan empat bahasa: Arab, Melayu, Sunda, Jawa [= Risalah Qiblat described in four languages: Arabic, Malay, Sunda, Javanese]. https://archive.org/details/RisalahQiblat4Bahasa


