15

Development of A Unicode Compliant Kashmiri

Font: Issues and Resolution

Aadil A. Lawey
Nazima Mehdi

Introduction: Kashmiri language is primarily spoken in the Kashmir province and some parts of the Jammu province of the state of Jammu and Kashmir and by migrant populations in the rest of India and abroad. The earliest script used for writing Kashmiri was the Sharda script which is now only used by some Kashmiri pundits for writing horoscopes. Presently, the official script of Kashmiri is the modified Persio-Arabic script with additional diacritic marks to represent Kashmiri specific sounds. Alternate scripts like the modified Devanagari and Roman script are also used for writing Kashmiri. Regarding the modified Persio-Arabic script, it is written from right to left, and it has two modes: *nasakh* or the type script, and *nastalikh*, the handwritten version.

Issues Observed:

Same Character Having Different Shapes at Different Positions: Like the typical Persian and Arabic alphabets, while writing Kashmiri using the modified Persio-Arabic

script, letters change their shapes. The shapes vary with position i.e. at isolated, initial, medial and final positions.

One example is that of "ه" 'heh' which is pronounced as 'h' have different shapes at different positions at the unit's initial, medial and final position.

Isolated form	ه
Initial form	ﻫ
Medial form	ﻬ
Final form	ﺢ

So in order to solve such an issue different glyphs were needed for a single character to represent its shapes at different positions.

Different Characters Having Same Sound: Another issue was the redundancy of characters e.g. the characters ظ, ض, ز, ذ (though differently pronounced in their language of origin) have no different phonological forms for a Kashmiri native speaker. They are all pronounced as 'Z' but are represented by four different characters. For such cases different characters needed to be developed though they represent the same sound.

Unicode Issues: No unicode for Kashmiri specific characters like 'palatilized Yeh', 'medial palatilization' and 'wavy hamza below'. Also the shape of 'dodI va:v' doesn't resemble with Kashmiri dodI va:v.

ؼ	**No Unicode**
ﻟ	**No Unicode**
ؚ	**No Unicode**
ﻭ	**Incorrect Shape**

Tentative Solution:
Kashmiri Characters with Unicode Values: The first step undertaken was to take the Unicode values of all the characters of Kashmiri which were available in the Arabic code chart. These are given as under.

Letters	Unicode Value	Letters	Unicode Value
ا	0627	ک	06A9
ب	0628	گ	06AF

پ	067E	ل	0644
ت	062A	م	0645
ٹ	0679	ن	0646
ث	062B	و	0648
ج	062C	ہ	06C1
چ	0686	ھ	06BE
ح	062D	ء	0621
خ	062E	ی	06CC
د	062F	ے	06D2
ڈ	0688	آ	0622
ذ	0630	أ	0672
ر	0631	؟	061F
ڑ	0691	؛	061B
ز	0632	-	06D4
ژ	0698	ِ	0650
س	0633	َ	064E
ش	0634	ُ	064F
ص	0635	ٗ	0657
ض	0636	ٚ	065A
ط	0637	ٛ	065B
ظ	0638	ٔ	0654
ع	0639	ٕ	0655
غ	063A	ٖ	0656
ف	0641	ٍ	064D
ق	0642		

Table 1.1 Kashmiri characters with Unicode Value

Regarding the other characters of Kashmiri which were not present in the Arabic code chart, appropriate design tools were used to create the necessary characters and glyphs. The tentative solution for Unicode which was employed was to assign temporary Unicode values of Arabic code chart to those Kashmiri characters. These new characters were assigned Unicode values of those Arabic characters which:

- were not present in Kashmiri.
- behaved in the same way in the initial, medial, final and isolated positions as required in Kashmiri.

Characters which were Developed are:

medial palatalization   ڈ

palatalized yeh   ﯼ

dodI vaav   ۏ

wavy hamza below   ۭ

Medial palatization (ڈ) was formed by using two Unicode characters 06EA (Arabic Empty Centre Low Stop ۪) and 066E (Arabic Dotless BEH ٮ).In this way medial palatalization issue was solved and it is represented by two Unicode values as given above. Another issue was regarding the shape of "dodI va:v"(ۏ which was solved by changing the shape of glyph "va:v with the ring"(ۏ).

The Temporary Unicode values assigned to these Characters:

Developed Characters	Temp. Unicode Values
ۭ	06ED
ﯼ	06CD

Table 1.2 Charcters having Temporary Unicode Values

In the recent Unicode meeting the two Kashmiri specific characters palatalized yeh(ﯼ)and wavy hamza below(ۭ got the approval for Unicode values, which can be substituted for the temporary values and any corpus which is produced here can be saved from any legacy issues.

Same Character Having Different Shapes at Different Positions: As same characters change shapes at different positions e.g. "ہ".This issue was handled by having different glyphs for that particular character to represent its shapes at different positions. Glyph substitution was handled with the use of glyph substitution table (GSUB) and glyph positioning was handled with the use of glyph positioning table (GPOS).

Font Developed: The font '*Afan Koshur Naksh*' was developed by taking into consideration the issues and their solutions as discussed above. The font is completely workable and has been tested by people who are professionally involved with Kashmiri teaching and typing to test it for us and the response is very good.
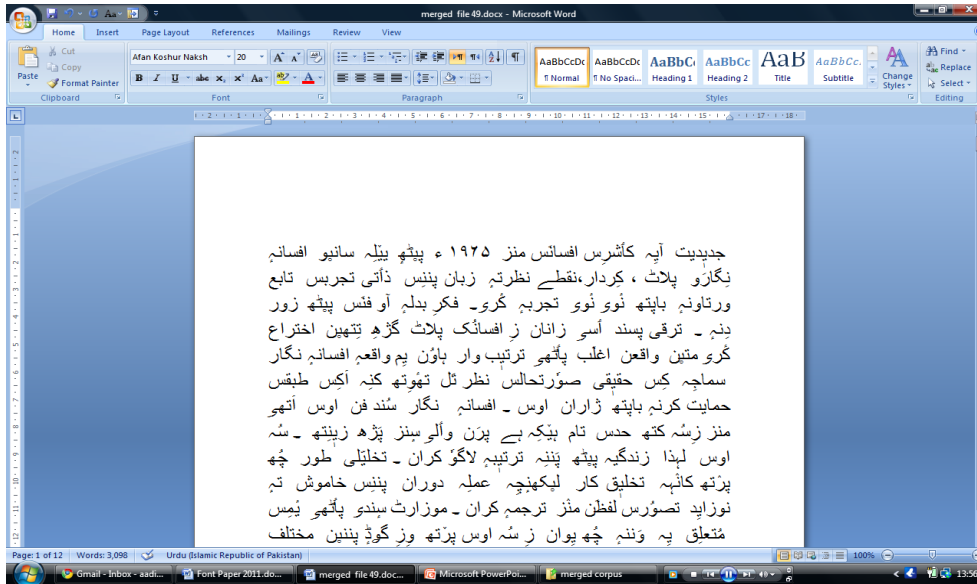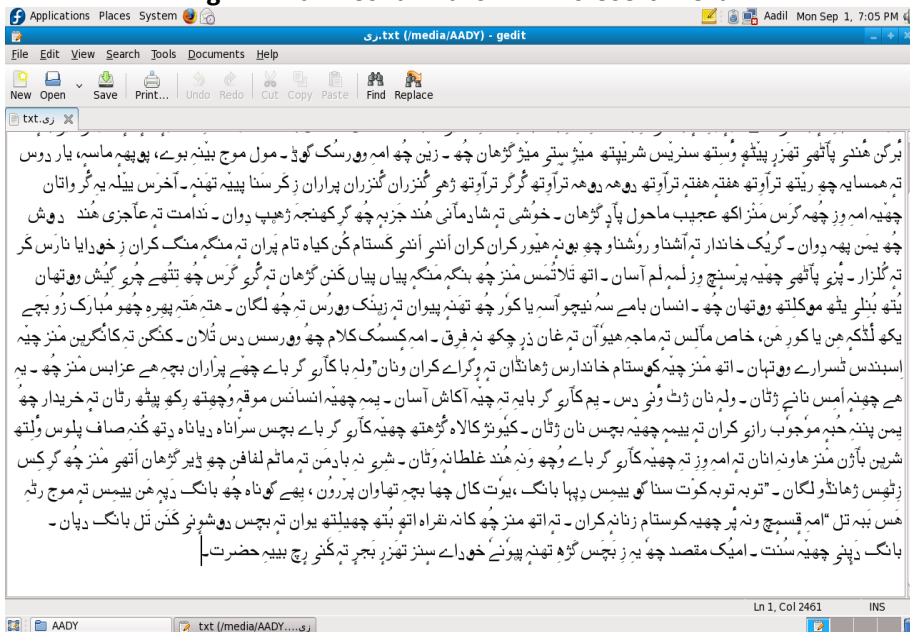
Fig 1.1 Afan Koshur Naksh in Microsoft Word

Fig 1.1 Afan Koshur Naksh in gEdit Text Editor (LINUX)

Conclusion: The font created (AFAN Koshur Naksh) so far is the first Unicode based font for Kashmiri. It is the first step towards the NLP framework of Kashmiri, as without the font it would not have been possible to build NLP tools like Synset creation, Dictionary, Morph Analyzer, POS Tagger and so on. The font created is comptabile with both windows and Linux and is being used for creating an e-corpus for Kashmiri and approximately a corpus of around 4 lakh words is developed.

References

Lupton, E. 2004.*Thinking with type: A critical guide for designers, writers, editors, & students.* Princeton: Princeton Architectural Press.

Macmillan, N. 2006. *An A–Z of type designers.* Yale: Yale University Press.

Online Etymology Dictionary.

http://dictionary.reference.com/browse/fonthttp://www.unicode .org.

Webcites:

Unicode Arabic Chart page. http://www.unicode.org/charts/PDF/Unicode-3.2/U32-0600.pdf

http://en.wikipedia.org/wiki/Font acessed on 15-feb-2011

MICROSOFT VOLT 1.3 Release Notes.

http://www.microsoft.com/typography/VOLT%20Release%20Notes/default13.mspx accessed on 11-feb-2011

http://www.unicode .org.

❑❑❑