Center of Excellence for Urdu Informatics
National Language Authority
H-8/4 Islamabad

Dear Jonathan!

I received your comments regarding our proposal for addition of *Preliminary Proposal to add Nuqta Characters to Arabic Block* (L2/06-039 now moved to L2-06-240). Reposing our stance regarding the *Nuqtas*, These characters have individualistic script existence and are often needed in the generation of electronic texts like pedagogical material. You will find scanned images of two primers published by the Pakistani provincial governments in wide circulation which use *Nuqtas* as separate characters. In situations like these, we have to resort to the old fashioned way that is the calligrapher leaving our DTP systems of no use. And definitely this introduces much more complications in the electronic storage, retrieval and transmission of these texts. Previously, Unicode added many entries from our ASCII Code Plate notification including the notion of ghost characters thus completing the set of ghost characters of the Arabic script. Now it's complimentary to add support for these Nuqta characters to the Arabic code block to realize the real benefit of the set of ghost characters.

Moreover, *Nuqtas* are also present in *Quran* as separate characters like 2, 3 and 4 *Nuqtas* above used separately. In these circumstances, need for these *Nuqta* marks as separate characters is of immense importance.

Regarding the current encoding model followed by Unicode about Arabic script, we do not intend to change the system or introduce a parallel or duplicate encoding system in the Arabic Block. Merely, what we want is the addition of these *Nuqta* characters along with the proposed properties. If it introduces a parallel system then it's an additional benefit yielding self sufficiency of the Arabic script.

However, a point that we would like to make is that enormous numbers of languages are adopting Arabic Script as the script of their choice and are defining their own unique ways to represents sounds not available in the Arabic phonetic system. Usually, registering a character to Unicode is a lengthy process spanning over several years and this waiting queue merely results in take-off delay of these languages into the digital age. But the presence of these proposed characters will provide somewhat "Grand Unification Theory" like effect to the Arabic script. Regarding your mention about costs, definitely every transition has associated costs involved but the associated benefits must be considered as well. Arabic script has three major languages: Arabic itself, Urdu and Persian. Currently, the amount of electronic text in these languages is not that much enormous as compared to the languages written in Latin script like English French etc. So a radical change at this stage is invite able otherwise delay will further make it impossible to impose any restructuring on the Arabic Block thus yielding a wrong direction.

Definitely, the idea of combining marks generating new characters is not new to the Unicode. Many existing scripts covered by the Unicode enjoy the support for combining diacritics marks that sit of the base characters in order to form new characters. Latin is just an example of that (See Section 7.7 from The Unicode Standard 4.0). Indic group of languages have also combining marks associated with them. Devanagri has such a code point (U+093C DEVANAGRI SIGN NUKTA) "for extending the alphabet to new characters" as per the documentation in the standard.

Current state of the Arabic Code block is somewhat analogous to the ASCII. ASCII was developed by IBM as their proprietary standard on their machines using 7 bits but it got wide acceptance getting an expansion of 8 bit including many characters from European languages. But later on, as the computing canvas expanded, ASCII was short enough to even support all the European languages and the reason was simple: ASCII was not considering the Latin Script rather just English (7 bits) and few European languages (8 bits). But later on, when the whole stage for the languages was considered,

decomposed marks were to be added. Same is the case with Arabic. The motive behind the initial digitization of Arabic was to capture the Middle East Markets. Thus vendors altogether ignored the rest of the languages of Arabic script having a narrow view about the script developing proprietary standards covering only Arabic and at most Persian. Arabic coverage of Unicode is simply reflection and continuation of that approach which is definitely inadequate for the future needs of the script

Lastly, regarding your mention about the font development using OpenType, we've already developed a *Nastaleeq* font (named Pak Nastaleeq) which stands as our National Standard supporting all the regional languages whether digitized or not and we've deployed the same techniques you described of decomposing at the glyph level. But to support this idea we need the addition of *Nuqta* marks into the Arabic Code block. This font is likely to be adopted by the Urdu versions of the Microsoft products replacing the traditional *Naskh* style screen layouts into *Nastaleeq* as we are in close negotiations with Microsoft as their Localization Program Partner regarding this.

The facts described herein can be rephrased into the FAQ mentioned in the Open Action Item 108-A2 designated to Lisa Moore.

Regards,
Dr. Attash Durrani
Liaison Officer
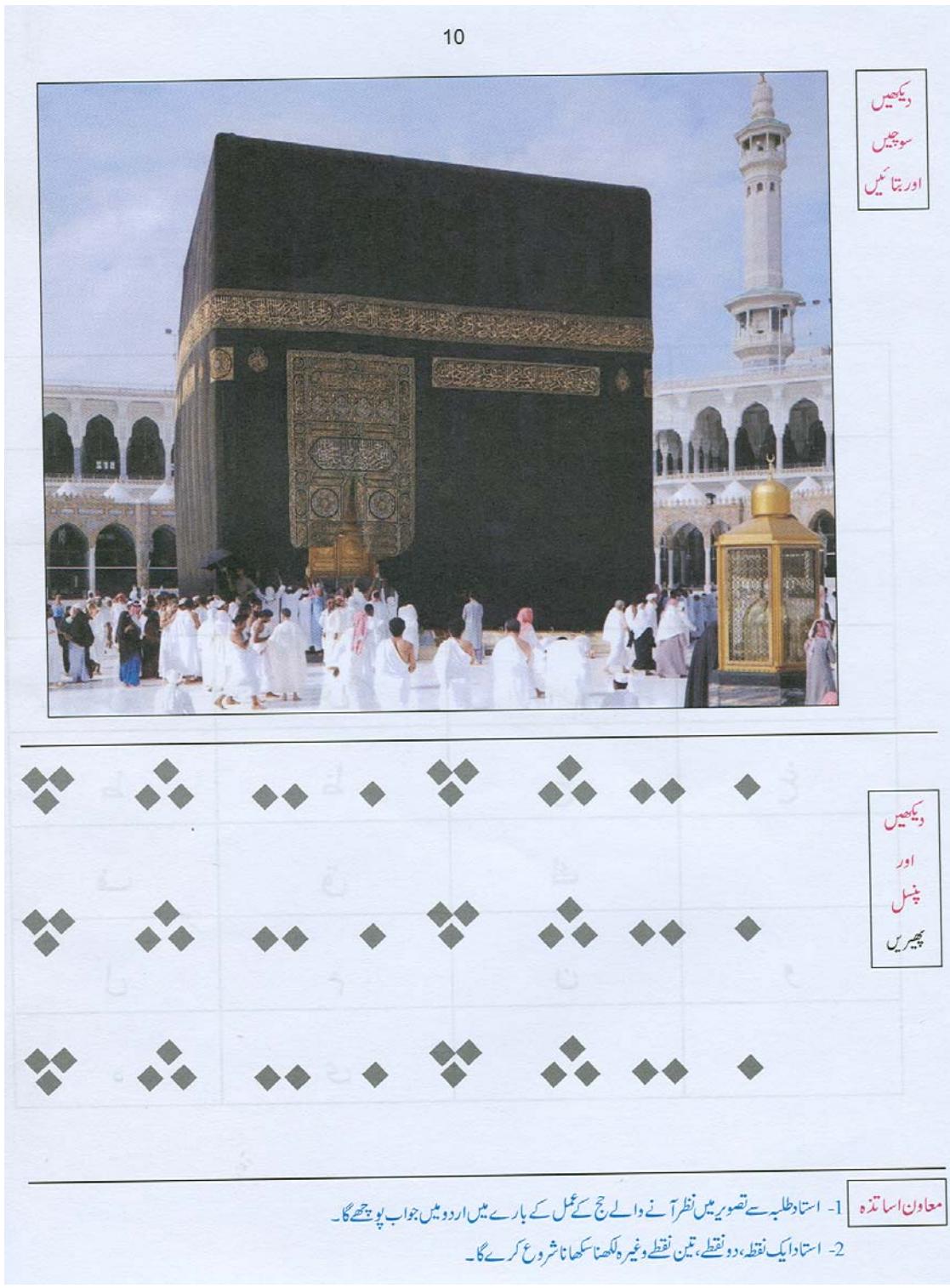National Language Authority
Pakistan

Fig 1: A Page from Primer published by Government of Punjab (Pakistan) depicting Nuqtas as separate characters
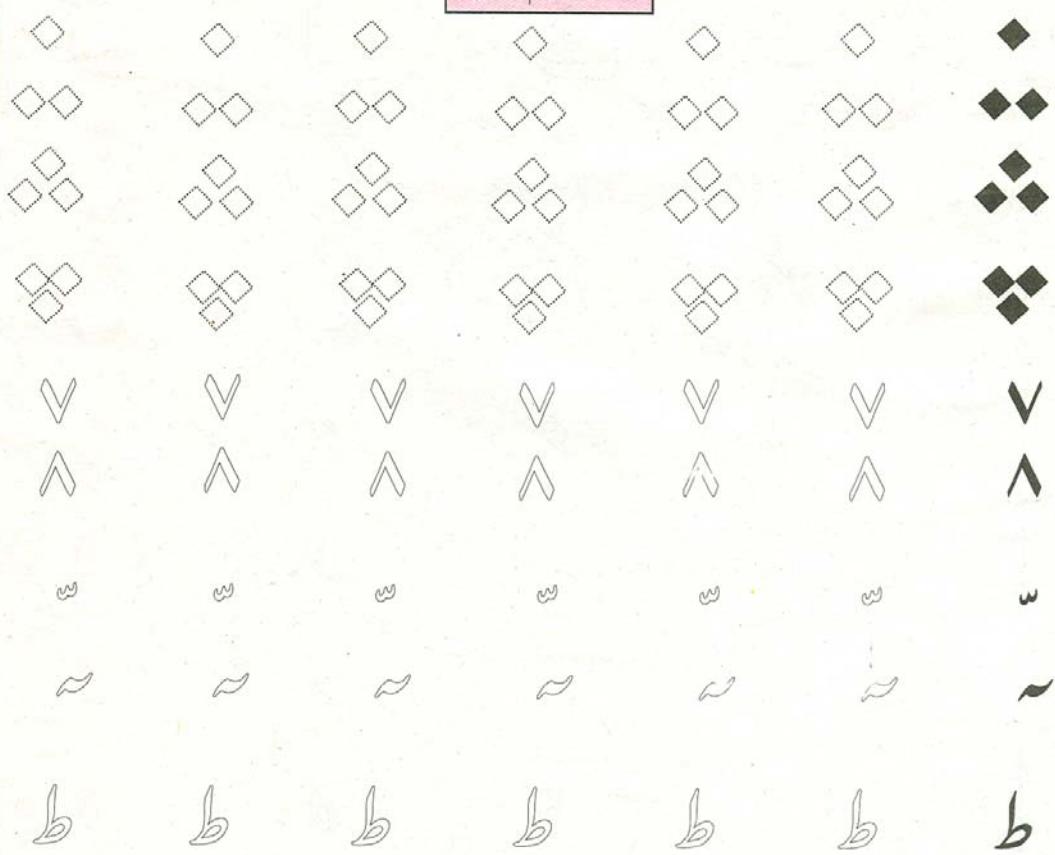
Fig 2: A Page from Primer published by Government of NWFP (Pakistan) depicting Nuqtas as separate characters