

## Explanation of Proposed Kazakh (in Arabic Script) Romanization Table

The proposed romanization of the Kazakh language in Arabic script fills a gap in the ALA-LC romanization tables. While the majority of current Kazakh publications are in Cyrillic, a number of titles are issued in China using a modified Arabic script. Some of these titles are acquired by U.S. libraries and cataloged in OCLC, and the lack of an approved romanization table has led to problems, for example, duplicate records and searching difficulties. The University of Washington has a small collection of Kazakh books we plan to catalog, so we have made this proposal in the hope of standardizing the romanization before we process them.

In constructing the table, we followed the *Procedural Guidelines for Proposed or Revised Romanization Tables* ([http://www.loc.gov/catdir/cpsd/romguid\\_2010.html](http://www.loc.gov/catdir/cpsd/romguid_2010.html)), hereafter called the ALA-LC guidelines.

### Sources of Information

We drew the alphabetic order and initial/medial/final forms of letters from alphabet books that give contemporary usage: *Alippe : besjildiq bastawış mektep oqwliǵı* (1987) and *Alippe : altijildiq bastawış mektep oqwliǵı* (1994). We added an additional mark because Kazakh in Arabic script uses a high *hamzah* to indicate front vowel harmony when it is not apparent from implicit factors. The hamzah must be represented in the romanization in order to have full reversibility.

We chose to present the letters of the alphabet in a single sequence and not to separate the vowels, as is often done in romanization tables for languages in Arabic script. First, the ALA-LC guidelines deemphasize pronunciation, so it seemed less important to focus on vowels. Second, Kazakh in Arabic script has letters that are difficult to characterize as consonant or vowel, e.g. the diphthong *ǵ*, which may be pronounced as either “u” or “w” depending on context. The *hamzah* is also neither a consonant nor a vowel.

### Romanization Schemes Considered

We reviewed a number of romanization schemes that render Kazakh in Latin letters. They are listed here with brief reasons why we found them unsuitable for a romanization based on ALA-LC guidelines. Background information and a good comparison of romanization schemes written in English can be found in “Kazakh Roman Alphabet Perspectives” ([http://www.eki.ee/wgrs/wgr06\\_9.htm](http://www.eki.ee/wgrs/wgr06_9.htm)). A fuller article in Russian is available in the Russian Wikipedia ([http://ru.wikipedia.org/wiki/%D0%9A%D0%B0%D0%B7%D0%B0%D1%85%D1%81%D0%BA%D0%B8%D0%B9\\_%D0%B0%D0%BB%D1%84%D0%B0%D0%B2%D0%B8](http://ru.wikipedia.org/wiki/%D0%9A%D0%B0%D0%B7%D0%B0%D1%85%D1%81%D0%BA%D0%B8%D0%B9_%D0%B0%D0%BB%D1%84%D0%B0%D0%B2%D0%B8)

[http://en.wikipedia.org/wiki/BGN/PCGN\\_romanization\\_of\\_Kazakh](http://en.wikipedia.org/wiki/BGN/PCGN_romanization_of_Kazakh)), which has useful tables. The BGN/PCGN table is available at: [http://en.wikipedia.org/wiki/BGN/PCGN\\_romanization\\_of\\_Kazakh](http://en.wikipedia.org/wiki/BGN/PCGN_romanization_of_Kazakh).

BGN/PCGN romanization uses blends instead of single letters and does not conform to guideline III.3.a

Russian GOST Roman A uses unusual diacritics and diacritics below letters, which does not conform to guidelines III.3.b-d

Russian GOST Roman B uses blends instead of single letters and does not conform to guideline III.3.a

Simplified Pinyin is used by Kazakhs in China but, depending on the version cited, has non-standard characters and blends or unusual diacritics, which does not conform to guidelines III.3.a-c

International Phonetic Alphabet is focused on pronunciation and uses special Latin characters, which does not conform to guideline III.2.c

In addition to reviewing existing schemes, we constructed a romanization that gives the same result as the romanization of Kazakh from Cyrillic, but because vowel harmony is incompletely indicated in the Arabic script, the rules of application are so complex that the romanization does not meet the goals of reversibility and machine transliteration.

The romanization for Uighur in Arabic script is not suitable for Kazakh. First, there are letters in Kazakh that do not exist in Uighur, for example  $\text{ڤ}$ . Second, the same symbol in Uighur and Kazakh may represent different letters, for example  $\text{ا}$ , which is  $\text{ä}$  in Uighur but  $\text{e}$  in Kazakh (the equivalent of  $\text{ä}$  in Kazakh is the letter  $\text{ӱ}$ ).

We could not find an official romanization used by Kazakhs in China.

### **Proposed Romanization**

In the end, we settled on a modified version of a romanization by the Kazakh National Information Agency (known as QazAqparat or Kazinform). In 2004 QazAqparat put forward a romanization, which was considered as part of a plan to switch from Cyrillic to a Latin alphabet in Kazakhstan. This romanization does not have official status because the plan was put on hold in 2006. Based on the Turkish alphabet, the QazAqparat romanization uses single letters and standard diacritic marks. Its problematic areas are diacritics below two letters and the undotted  $\text{ı}$ .

In the proposed table, we adopted the QazAqparat scheme with the following changes:

The cedilla below the letters  $\text{c}$  and  $\text{s}$  was replaced with a hacek;

The undotted *ı* was replaced with a dotted *i*. Two further changes were then required: the corresponding front vowel was changed from dotted *i* to *ï* (which is consistent with the treatment of other vowels in this scheme), and the *ï* was changed to *y*;

The alif was added as the romanization of the high *hamzah* when needed.

Note that Kazakh in Arabic script does not make use of several letters that are distinctly Cyrillic: *ц, щ, э, ю, я, ъ, and ь*. In limited cases where some of these sounds are needed, they are represented by a combination of Arabic letters and may be romanized from those letters in a straightforward way. Also, two separate letters in the Cyrillic alphabet (*и* and *й*) are given by the same Arabic letter (*ي*), so the letter *y* in romanization works better in words where it corresponds to *й* (e.g. Ara. *aytqanda* = Cyr. *aïtkanda*) than in words where it corresponds to *и* (e.g. Ara. *qysa* = Cyr. *qissa*). Another case where the romanization of a letter works better in some cases than in others is the diphthong *ڭ*, romanized as *w*: it looks better when it occurs between vowels (e.g. Ara. *bastawiš* = Cyr. *bastauysh*) than when it occurs between consonants (e.g. Ara. *oqwliq* = Cyr. *oqulyk*). Finally, vowel harmony for front vowels is not marked in medial or final forms, so the corresponding cells of the table are blank.

### **Alternatives to the Proposal**

We note here some of the obvious alternatives to the choices we have made in the proposed table.

Instead of introducing changes to the QazAqparat scheme, it could be adopted as it stands, with the single addition of the *hamzah* for reversibility. On the positive side, this would promote compatibility with what may become the Latin alphabet of Kazakhstan. On the negative side, it would not conform to guidelines III.2.c and III.3.d.

In terms of single letters, Arabic *ي* could be romanized, instead of *y*, as the letter *i* plus a diacritic such as the breve, which would work better in many words (e.g. the awkward *Šynjyañ* would become *Šïnjïañ*). Also, Arabic *ڭ* could be romanized, instead of *w*, as the letter *u* plus a diacritic such as the macron (e.g. *oqwliq* would become *oqũliq*).

### **Unicode and Representation of Arabic Letters**

Unicode offers a variety of code points to represent Arabic characters. We chose presentation forms whenever available, and have provided a separate table with the Unicode values used in the romanization table. You may decide on other code points for a final version.

We are not satisfied with the display of Arabic letters in the Arial font in Microsoft Word, in particular for medial forms. While the letters link correctly with surrounding letters in

continuous text, the medial forms often look like initial forms when standing alone in the table. In a final version, you may have a different font that does a better job of representing these letters.