An Introduction to Latin-Script Uyghur

by
Waris Abdukerim Janbaz
Paragraphe Laboratory
Researcher & PhD student**,** University of Paris VIII
warisabdukerim@yahoo.com

Prof. Imad Saleh
Paragraphe Laboratory
University of Paris VIII, idsaleh@orange.fr

Dr. Jean Rahman Duval
rahmanjean@yahoo.com

Abstract: In this paper, we briefly review the failure of the shift of Uyghur from its original Arabic script to the Pinyin-based "New Uyghur Script" then discuss the development of a new standard — Latin-Script Uyghur. We also explain its orthographic rules, present its status and consider future perspectives.

Introduction

Uyghurs are a Turkic-speaking ethnic group, said to be descended from the historical Huns[1]. Their population, estimated at about ten million, is distributed over a Central Asian region including today's Xinjiang Uyghur Autonomous Region (XUAR) of China (also sometimes referred to as Chinese or Eastern Turkistan), parts of Kazakhstan and urban regions in the Ferghana valley. Historical records indicate that Uyghurs have used various scripts, sometimes concurrently, during different periods over the past 2000 years. The most important ones are the Soghdian Script (5th -10th centuries[2] and 15th-16th centuries [3]), the Orkhon-Yenisei script (also known as Turkic Runic script, 6th-9th centuries)[4], the Uyghur Script (also known as Old Uyghur Script, 10th-18th centuries)[5] and,

---

[1] Sima Qian, _Shiji _(Historical Records_)_, *Xiongnu zhuan* (the Section on Huns), Volume 9; Abdüshkür Muhemmed Imin, _Qedimki Merkiziy Asiya_ (Ancient Central Asia), Xinjiang People's Press, 2002, pp. 95—110. This claim may be partly valid for present-day Uyghurs, who have kept the ethnonym of the historical Uyghurs but whose ethnogenesis reflects the mixing of populations of diverse origins through centuries of migrations and invasions (for a discussion of that ethnogenesis see J. R. Duval: "Modern Uyghur, A Historical Perspective" in: Culture Contact, History and Ethnicity in Inner Asia, № 2. Michael Gervers and Wayne Schlepp, eds. Toronto: Joint Center for Asia Pacific Studies, 1996, pp. 132-67.).

[2] Haji Nurhaji, *Qedimki Uyghurlar we Qaraxaniylar* (*Ancient Uyghurs and Qarakhanids_)_, Xinjiang People's Press, 2001, p.122.

[3] Abduréhim Rahman & Ömerjan, _Shinjang Tarix Matériyalliri_ (*Xinjiang Historical Documents_)_, volume 40, Xinjiang People's Press, 1980.9.

[4] The Orkhon alphabet is believed to be derived from a non-cursive Soghdian script. It was used by the Orkhon Uyghur Empire (744 - 840) but also by other Turkic tribes of Eastern Central Asia. See

after their gradual adoption of Islam starting in A.D. 934, the Arabic Script (10th century - present).

Modified to take into account the phonetic and grammatical characteristics of Turkic languages, the Arabic script (sometimes referred to as *Chaghatay* script) was used continuously by Uyghurs and other Central Asian Turkic peoples until the 1920s. Several efforts were then made to reform it. Besides spelling changes, these also included at least three attempts to replace it with the Latin or Cyrillic scripts. Many of these changes were politically motivated. Most of the early endeavors were directly linked to Stalin's Nationality Policy, embodied by the decision made at the First International Turkology Congress, held in Baku in 1926, to Romanize the newly created "literary languages" of Central Asian Turkic languages. The subsequent shift of the Republic of Turkey to the Latin alphabet in 1927-1928 was instrumental in bringing about the 1937 decision by the Soviet Union, concerned by a potential "pan-Turkic" threat, to impose Cyrillic-based alphabets to its Turkic peoples[6].

China, where most Uyghurs live, became particularly active in language and script reform after the 1949 Communist victory, which led to the adoption of many Soviet-style approaches to language policy. At the time, Russian linguists were quite active in helping China repertory and codify Chinese "minority"(non-Han Chinese) languages. In this context, Uyghur dialects, spoken on both sides of the China-USSR border, were unified into a single "literary language" based on the Soviet standard. Further, in March 1956 the State Council of the People's Republic of China (PRC) and the XUAR Language and Script Committee (*Shinjang Uyghur Aptonom Rayonluq Til-Yéziq Komitéti*)[7] (hereinafter XLSC) launched a short-lived campaign to impose the Cyrillic script[8] among Uyghurs, Kazakhs, Kirghiz, Mongols and Sibes (*xībó*) living on its territory.

When relations with the USSR entered a period of strong tension, China decided to backtrack on linguistic policies that created too many links between Uyghurs (and other Central Asian peoples) who bestraddled the border, In November 1959, it promulgated a draft proposal for a Chinese "*Pinyin*"-based New Script (*yéngi yéziq* or "New Script" for Uyghur, hereinafter PNSU) which was adopted[9] for Uyghur and Kazakh at the first National Linguistic and Orthographic Conference of the XUAR, approved by the XUAR's People's Congress in the same year. Between 1960 and 1964, PNSU was

---

http://www.omniglot.com/writing/orkhon.htm

[5] Adopted, in an adapted form, first by the Mongols, and then by the Manchus.

[6] See Kamal Talibzade, *Learning to Read All Over Again: Alphabet Changes in Azerbaijan Throughout the Century, An Eyewitness Account*, Azerbaijan International 2000 (8.1), http://www.azer.com/ aiweb/categories/magazine/81_folder/81_articles/81_talibzade.html

[7] A XUAR government body created for the purpose of "fostering the development of Minority (i.e. non-Han Chinese) languages and writing systems in the XUAR".

[8] This modified Cyrillic script was, for Uyghur, borrowed directly from that which had been imposed on Soviet Uyghurs in 1947 to replace the original Romanized script decided upon in 1928.

[9] See "*wéiwúěr xīn wénzì fāng'àn — cǎoàn*" (*Draft proposal for the Uyghur New Script — Preliminary*) 《维吾尔新文字方案》(草案) and "*hāsàkè xīn wénzì fāng' àn — cǎoàn*" (*Draft proposal for Kazakh New Script — Preliminary*) 《哈萨克新文字方案》(草案), November 1959, http://www.xjyw.gov.cn/han/wenzi_gai.htm

implemented on a pilot basis, and then promulgated all over the XUAR. It became the only authorized script for Uyghur and Kazakh until the Arabic Script was officially reinstated in September1982 (XUAR Government Document No. XH-1982-283), after further modifications, but it kept the status of authorized option[10].

Background

The creation and promulgation of PNSU has also to be viewed against the overall linguistic environment of the PRC at the time. In an effort to eradicate illiteracy, the Beijing leadership was working towards the eventual replacement of Chinese characters by a Romanized script (*Hanyu pinyin*, or simply *pinyin*) and the script reform of "minority" languages had to be consistent with an overall plan involving a common script for all the ethnic groups of China.

The eventual failure of this attempt to Romanize Uyghur (and other XUAR ethnic writing systems) can be ascribed to the following factors:

Cultural resistance: the Arabic script represented a link with the religious tradition (Islam) and the cultural past for many Uyghurs who were reluctant to part with their identity and resented the reform as one more step towards assimilation and a loss of their linguistic autonomy.

A feeling of estrangement from other Turkic peoples: Turkey and other Turkic communities had been undergoing script reforms that had created a variety of different transcriptions for similar sounds. China, in its Romanization scheme, was going one step further by introducing unique new letters (ө, ŋ, ə, z, қ and ⱨ[11]) which isolated Uyghur from other Turkic languages. The use, adapted from Chinese *pinyin*, of the letters *q* and *x* for the phonemes [tʃ] and [ʃ] reinforced that feeling[12].

A fear of linguistic assimilation: the introduction of the letter groups *ch*, *sh* and *zh* used to transcribe purely Chinese phonemes made it possible to import Chinese words directly into Uyghur. This was perceived as yet another way of speeding up the sinization of the language[13]. Interestingly, PNSU did not go all the way in imposing Chinese pinyin

---

[10] On language and writing reform, see J. R. Duval: "Modern Uyghur, A Historical Perspective" in: Culture Contact, History and Ethnicity in Inner Asia, № 2. Michael Gervers and Wayne Schlepp, eds. Toronto: Joint Center for Asia Pacific Studies, 1996, pp. 132-67.

[11] Each of ө, ə, and қ was borrowed from the Cyrillic alphabet modified for Central Asian languages. z, ⱨ and ŋ were PRC creations.

[12] In respect of the letters *q* and x, and the letter *j*, it is interesting to note that they do not transcribe in PNSU exactly the same phonemes as in *pinyin*, where they correspond respectively to [tɕʰ], [ɕ] and [tɕ], and not to [tʃ] [ʃ] and [dʒ]. For non-Chinese speaking Uyghurs, however, there was no difference in pronunciation. See http://en.wikipedia.org/wiki/Pinyin

[13] Many new words, particularly those of a political nature, were imported directly from Chinese into Uyghur during that period, which corresponds to the Great Proletarian Cultural Revolution and a series of campaigns against Uyghur nationalism. Most of them disappeared to be replaced with "native" terms as

spelling since Chinese words imported into Uyghur still had to undergo certain changes, such as the substitution in PNSU of *y* to *pinyin i* in diphthongs *ai* and *ei* and that of *ə* to *pinyin e* or *a* in certain phonetic environments, thus undermining somewhat the purpose of linguistic unification.

Reform fatigue: in a short period, Uyghur had undergone a series of script reforms (from modifications in the Arabic to a changeover to Cyrillic and now to the Latin script). In the meantime, ASU had also undergone changes that made it much easier to use for a non-Arabic language[14], thus greatly reducing the need to replace it.

Since millions of XUAR Uyghurs had been educated in PNSU, but not in ASU[15], for the nearly twenty years between 1965 and1982, they became illiterate after the resurrection of ASU as the official script. Meanwhile, CSU[16] remained the only script used by Soviet Uyghurs[17] (as is still the case).

This chaotic situation created obstacles in the written communication of Uyghurs across geopolitical as well as generational boundaries. The advent of new communication technologies, far from solving the problem, further amplified it.

In the past two decades, computers and modern technology equipment have brought about new hopes for information processing and communication. Unfortunately, lack of script unification and the existence of a variety of unique characters in PNSU, CSU and ASU slowed progress in the penetration of information technologies. Lack of complete sets of ASU, CSU and PNSU standards, both in ASCII and Unicode, forced Uyghur computer users to resort to a wide assortment of non-normalized transliterations to share and process information in their native tongue (or to other languages—Chinese, Russian and English mostly). On one hand this contributed significantly to the familiarization of Uyghur computer users with the Latin alphabet and accelerated the demise of PNSU due to the presence in it of letters that are absent from most computer systems. On the other hand, it caused the multiplication of mutually incompatible Romanization schemes on information platforms.

By 2000, according to a survey the authors made, there were at least eighteen different but actively used Latin alphabets for Uyghur. This number does not include the unsystematic transliterations that cropped up everywhere due to the absence of a common

---

soon as the Arabic script was reinstated (see J. R. Duval, *op.cit*. note 9).

[14] Arabic letters used for purely Arabic phonemes (ث , ذ, ح, ط, ظ, ص, ض, ع) were gradually dropped from the 1920's on, while others (namely ا, ه , و and ي), in increasingly modified forms (addition of a *hamzeh on a tooth* to replace the initial silent *alif* in front of initial vowels; creation of an initial ئه [e]; gradual specialization of و into first ئو [o, ö] and ئۇ [u, ü] then further into ئو [o], ئۆ[ö], ئۇ [u] and ئۈ [ü]; and creation of the letterئی [i] differentiated from ي, and later supplemented with ئې [é]) to transcribe vowel sounds absent from the traditional script. Through these modifications ASU had become a phonemic script with one letter for every distinct sound. At the same time, it had become an independent alphabet which no longer followed the traditional rules of the Arabic alphabet.

[15] See http://www.omniglot.com/writing/uyghur.htm

[16] See http://www.omniglot.com/writing/uyghur.htm

[17] See annex 1

usable system. Some of these have a name (Angel, Kiboon, Makanim, ML, Hashim, PBLSU, UIY [*Uyghur Internet Writing*], Zaman) and some do not. Scholars proposed different alphabets with the "flavor" or influence of the country in which they lived or of the foreign language they spoke. It is easy to detect, among the eighteen alphabets, Turkish, German, Russian, English, and Chinese influences or various creative elements imposed on the alphabets.

Between November 2000 and July 2001, five conferences[18] were held at Xinjiang University in Ürümchi to introduce a unified LSU alphabet[19]. Among the attendees there were Uyghur linguists like Ibrahim Mutihi, Mirsultan Osman, Muhebbet Qasim, Imin Tursun, Abliz Yaqup, Xemit Zakir, Arslan Abdulla as well as computer experts such as Hoshur Islam,Türgün Ibrahim and young postgraduate students, including the second author. For obvious political and cultural reasons, the conferences strongly emphasized that "the proposed alphabet should not replace ASU nor should its introduction represent a new reform of the writing system. It is to be used solely in computer-related fields as an ancillary writing system".

In order to alleviate fears of yet another writing system reform and emphasize the limited scope of the project, it was decided to call this Romanization scheme "*Uyghur Kompyutér Yéziqi*"(UKY) or *Uyghur Computer Script*. This name was changed by the Uyghur Computer Science Association three years after to *Uyghur Latin Yéziqi*"(ULY), or *Latin Script Uyghur* (LSU), since "computer script" was perceived as not making much sense and could lead to a confusion with binary — 1 and 0.

The LSU project was covered by the official XUAR media and on the internet to inform the public of the effort. The media, in particular, was very careful not to send the wrong signal of an incipient writing reform. Nevertheless, even today some people still hesitate to use the term ULY (*LSU*) since they fear its potential association with an attempt to reform the common script. Others think it is important to have one-to-one correspondence (or a norm) between LSU and ASU. There are also people who believe that Uyghur should eventually follow in the tracks of Turkish, Azerbaijani and Uzbek and adopt a modern Latin-based writing system (similar plans have been made but not yet implemented in other Turkic republics, such as Kazakhstan, Kirghizstan and Turkmenistan), thus going against official state policies.

A report was sent to the XUAR government authorities in the spring of 2001 and, as a response, the XUAR People's Committee organized a hearing which recognized the value of the project and recommended that large-scale pilots be conducted to evaluate its feasibility.

Basic principles followed in the creation of LSU

Significantly, after multiple reforms, the ASU alphabet itself had evolved. It now

---

[18] See http://www.ukij.org/teshwiq/UKY_Heqqide.htm for details.
[19] Originally called in Uyghur *Uyghur Kompyutér Yéziqi (UKY)*—Uyghur Computer Script*,* then *Uyghur Latin Yéziqi (ULY)*—Uyghur Latin Script. See www.ukij.org/teshwiq/UKY_Heqqide(KonaYeziq).htm

transcribes on a one-to-one basis all the phonemes of the Uyghur language although it is not a true phonetic transcription, since it reflects the basic or theoretical sounds of the language and not the multiple changes that occur in spoken Uyghur[20]. Since 2001 a certain number of spelling rules have been promulgated by the XLSC[21] to codify its use. In spite of successive reforms, the spelling rules still contain a number of inconsistencies, mainly with the final letters *b/p*, and *d/t* in words of foreign (mostly Arabic but also Russian) origin, where the official spelling hesitates between etymology and actual pronunciation[22].

LSU is a transcription of the official language of XUAR Uyghurs and follows the same spelling rules as ASU.

In this section, based on one of the author's participation in the Committee's work and on our documentation of its debates and decisions, we describe the four basic principles which guided the XLSC's work and led to the decisions that were made to create LSU.

A. **One-to-one correspondence between the LSU letters and the phonemes of the written language and, therefore, as much as possible, between the LSU, ASU and CSU letters[23].** The objectives here were to facilitate the learning of LSU by users of different script background and to avoid burdening them with yet another set of transcription/spelling rules.

B. **Absence of ambiguity and avoidance, as much as possible, of double letters.** This principle was to encourage the creation of a clear and logical transcription/transliteration system that would permit easy reading, prevent possible ambiguities (particularly in the case of combined vowels to express a single sound) and avoid the needless lengthening of words.

C. **Minimal use of diacritical marks, which should be restricted to very common ones.** At the time, most keyboards in the XUAR were of the North

---

[20] such, for example, as the pronunciation in most dialects of *n* as *m* in front of labials, the softening of *q* into [x] in front of another consonant, of *j* into [ʒ] in front of another consonant, the softening of final root letters *j* (→[ʒ]), *k*(→[g]), *p* (→[b]/[w]), *q* (→[γ]), and *t*(→[d]) in front of a suffix beginning with a vowel, the frequent dropping of the letters *r* and *l* in a post-vocalic position or the change of *l* into *m* before an *m*.

[21] Official spelling is set by the XLSC. Given the number of partial reforms official spelling has undergone in the past years and the impact of dialectal pronunciations, a number of variants of the same words can be found in written documents in the XUAR. In the case of CSU, spelling is usually based on the Northern Uyghur dialects (essentially characterized by the frequent substitution of the sound [ʒ] for *y* in the initial position) and tends to follow the actual pronunciation more closely, e.g. it drops the unpronounced final *t* after *s* or *sh* but reinserts it when a suffix causes it to be pronounced.

[22] E.g. three words coming from the same Arabic three-letter root k-t-b, كىتاب (kitab—book), كۇتۇپخانا~ (kutupxana—library) and مەكتەپ (mektep—school); ئىقتىساد (iqtisad—economy, pronounced ixtisat, from Arabic iqtisad) versus مەقسەت (meqset—aim, pronounced mexset, from Arabic maqsad). See *Hazirqi zaman Uyghur edebiy tilining teleppuz lughiti (Pronounciation Dictionary of the Modern Uyghur Dictionary Language)*, Nationalities Publishing House, Beijing 1988).

[23] For CSU letters, the one-to-one correspondence can be guaranteed only in texts following the XUAR spelling (see note 14).

American QWERTY type and few users were likely to install specific software for LSU on computers basically designed for English and Chinese-language input).

D. **Close correspondence of the chosen Latin letters with their common international phonetic value of the chosen Latin letters.** This principle was intended to ensure that letters be used in a way consistent with common international pronunciation, since to use them otherwise could confuse non-Uyghur speakers attempting to read, say, an Uyghur personal or geographical name [which English-, French-, German- or Spanish-speaker would know how to approximate the pronunciation of Ürümchi if it were spelled *Ürümqi* or *Vurvumqi or of Shemshi if it were spelled Xəmxi*].

The above principles were applied in a hierarchical order, with the priority always given to the first principle on the list but with due consideration for the other principles. The PNSU alphabet of the sixties and seventies did not pass the test of any of these principles: it had more letters or letter groups than ASU (and than there are phonemes in Uyghur), it was ambiguous since it proposed two transcriptions each (*q* and *ch*, *x* and *sh*, *j* and *zh*) for three Uyghur phonemes (Uyghurs pronounce Chinese sounds like their closest Uyghur cognates), it used diacritical marks and letters absent from the common set of Latin letters and it made Uyghur words thus written unreadable for whoever is not familiar with PNSU, thus imposing recourse to other types of transcription for Uyghur personal and geographical names outside of the XUAR. In Romanized maps of China, for example, most Uyghur places appear under their Chinese names in *pinyin* transcription—*Yining* for *Ghulja* and *Kashi* for *Qeshqer*. Uyghur personal names in ID documents are written using the pinyin transcription of their Chinese-language rendition—*Waresijiang Abudukelimu* for *Warisjan Abdukérim*. Nevertheless, it was difficult for people who had been trained in PNSU and become accustomed to the phonetic value of PNSU letters to adapt to a new transcription system. Besides, the semi-official status PNSU had retained was somewhat of a deterrent to change. Fortunately, the example of other Central Asian Turkic nations that were at the time moving to Latin-based scripts helped the project gather speed. On the basis of the above principles, a tentative standard was finally adopted by the conference attendees[24]. Below, we describe the details of decisions made for every letter.

Agreed-on letters

Among the variety of transcriptions in use, it was possible to identify the following 18 letters which were employed by all for the same phonemes. Since they corresponded to the above basic principles, they were readily accepted into the LSU alphabet by the conference committee.

ئا	ب	د	ف	گ	ئى	ك	ل	م
a	b	d	f	g	i	k	l	m
ن	ئو	پ	ر	س	ت	ئۇ	ي	ز
n	o	p	r	s	t	u	y	z

---

[24] See note 17.

Table 1. Agreed-on letters

Problem letters and decisions

For the other letters, there existed the following choices, based on the various transcription systems:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
چ	ch	ch	ch	ch	ch	ch	c	ç	ç	ç	ç	č	č	q	q	q	q	q, ch
ئە	e	e	e	a'	e	ä	e	e	e	é	ä	ä	ä	e	e	va	ae	ə
ئې	i	e	ë	e	e	e	e	é	é	e	e	e	e	e	e	e	e	e
غ	gh	gh	gh	gh	gh	gh	gh	ġ	ğ	ğ	ğ	ğ	ğ	v	w	vg	gh	ʁ
ھ	h	h	h	h	h	h	h	h	h	h	h	h	h	hf	wh	vh	v	ⱨ
ج	j	c	j	j	j	j	j	c	c	j	j	d_3	j	j	j	j	j	j, zh
ڭ	ng	ng	ng	ng	ng	ng	ng	ñ	ň	ņ	ň	ņ	ņ	nf	nh	ng	ng	ng
ئۆ	o	v	ö	o'	o	ö	o	ö	ö	ɵ	ö	ö	ö	oi	oe	vo	oe	ɵ
ق	q	q	q	q	q	q	q	ķ	q	ķ	q	q	q	c	k	vk	kh	κ
ش	sh	sh	sh	sh	sh	sh	x	ş	ş	ş	ş	š	š	x	x	x	x	x,sh
ئۇ	u	y	ü	u'	u	ü	u	ü	ü	ʉ	ü	ü	ü	ui	ue	vu	ue	ü
ۋ	v	w	v	v	v	w	w	v	v	v	v	v	v	w	w	v	w	w
خ	x	x	x	x	x	kh	h	h	ĥ	x	x	x	x	h	h	h	h	h
ژ	j	zx	zh	zh	j	zh	jz	ź	ź	ź	ź	$_3$	ž	zf	zh	vz	zj	ⱬ

Table 2. Problem letters and their proposed Latin equivalents

Choices had to be made following the basic principles. They are discussed below:

چ [tʃ]: principle C eliminated *ç* and *č*, and principle D removed *c* and *q*. ***Ch*** was chosen as it complied with all the rules, though it had the shortcoming of being a two-letter combination. It is read [tʃ] in most Latin-based writing systems of the world (as opposed to the PNSU *q,* Turkish *ç,* Slavic *č* or the simpler *c,* proposed in some systems) and it had been chosen for the same phoneme in Latin Uzbek. It is easily readable for Uyghurs trained in PNSU since it was used in that system for the phoneme [tʃ] in words of Chinese origin (e.g *gungchəndang*—Communist Party).

ئە [æ]: principle B eliminated *ə* and principle D eliminated *va.* The combined letters *ae* had the disadvantage of needlessly lengthening words as ئە is a very frequently used vowel (e.g. *maektaep* for مەكتەپ *mektep*—schoo*l).* The same was said for use of the apostrophe after an *a* which was proposed to comply with the Latin Uzbek use of an apostrophe to indicate a variant reading of the Latin letters *g* and *o* (school would have to be written *ma'kta'p*). Besides, the use of the apostrophe could have led to ambiguities, as we will see later. A number of people were in favour of chosing the letter *e.* But, given that [æ] is one of the three front vowels of Uyghur, many felt it would be best to use the letter *ä* to transcribe it, so as to have a systematic visual opposition between the back

vowels *a/o/u* and front vowels *ä/ö/ü*, as is the case in Hungarian and Finnish. They believed that *e* should be used for ئې, already transcribed as *e* in PNSU. Frequency considerations[25], however, led to the final adoption of the Turkish solution of using ***e***.

ئې [e]: since *e* was no longer available (see above), the choice was between *i, ë* and *é*. Principle A eliminated *i*. Eventually, é was chosen since the accent was felt to be easier to use than the diereses (a single diacritical mark instead of two). Besides, the sound is closer to the French *é* than to the Albanian *ë*.

غ [ɣ]: principles B, C and D eliminated *ğ, ġ, v, vg ,w* and *oɳ* and the letter *g* was already taken by ﯓ. The double letter ***gh*** was chosen over the Uzbek *g′*, as it is very common to represent غ in the transcription of words coming from languages in which the sound [ɣ] exists: e.g. Af**gh**anistan, Ma**gh**reb, A**gh**a Khan.

ھ [h]: principles B, C and D eliminated *v,* ꮒ*, vh, wh* and *hf*. The letter ***h*** was chosen as the simplest solution (no diacritical mark) and also because, in most languages, it represents a sound which is quite close to the Uyghur sound.

ج [dʒ]: principle B eliminated ʤ , principle D eliminated *c* and *zh*. The letter *c* is pronounced [dʒ] in Turkish and Azeri. For this sound, PNSU offered two choices, *j* in native words and *zh* in words of Chinese origin (e.g. *fangzhen*), the latter being used to represent that sound only in Chinese *pinyin. j* was chosen as closest to international usage.

ڭ [ŋ]: principle B eliminated *ñ, ň* and *ṇ* and principle D eliminated *nh* and *nf*. The choice naturally ended up being the compound letter ***ng***, which is commonly used in many languages to represent the sound [ŋ].

ئۆ [ø]: options *vo, oe* and *oi* were eliminated as they either needlessly lengthened words or did not render Uyghur sounds clearly for non Uyghur speakers. The letter *v*, chosen by some for its similitude with the top part of the ASU letter ئۆ was eliminated on the basis of principle D. The PNSU letter *ө* was eliminated on the basis of principle B. Given the need to differentiate with *o*, it was decided to use letter *ö* which, in spite of the inconvenience of a diacritical mark, is widely used in other languages to transcribe a sound that is similar or close to this Uyghur sound.

ق [q]: principles B and D eliminated *ḳ* and *ĸ*. Principle D removed *vk* and *c*. The letter *k* was already used for ك and the compound letter *kh* corresponds to a different phoneme in all other international systems. The letter ***q*** was chosen as it was the common transcription used for the same letter in Arabic-based writing systems (e.g. al-Qaida,

---

[25] according to Alim Ahat of Uighursoft Company (www.uighursoft.com), ك appears 67,121 times (4.065%) in a dictionary corpus of 1,651,130 letters, whereas ئې appears 13,037 times (0.790%).

Mohammed Iqbal, Iraq, Qatar). It is also the letter used for the same sound in Latin Uzbek and Tatar and for a cognate sound in Azeri.

ش [ʃ]: principles B and D eliminated *ş* and *š*. Principle D eliminated x, which is used to transcribe this phoneme in only few languages (mostly Chinese pinyin, Portuguese and Vietnamese). International, as well as Latin Uzbek use dictated the decision to choose *sh*. This decision was reinforced by the fact that, beside *x*, PNSU also proposes *sh* for words of Chinese origin such as *gungshe* (commune)[26].

ئۇ [y]: principles C and D eliminated *u', vu, ue* and *ui* and principle A eliminated *y* (an attractive choice since it is the IPA symbol for the front rounded high vowel) since it was already used for ي, and *u* which represents ئۇ. The letter *ü* was eventually chosen since it uses the same differentiating method with its back-vowel equivalent *u* as does *ö* versus *o*[27] and it is widely used around the world to represent a sound similar or close to the Uyghur sound in ئۇ.

ۋ [w]: this letter is used to render both a consonant wavering between a bilabial glide [w] and a labio-dental fricative, very reminiscent of the Dutch letter *w* (somewhere between a *w* and a *v*), at the beginning of a syllable (e.g. ۋەتەن [wætæn]—motherland, ۋەزىر [wæzir]—vizir, قولۋاق [qolwaq]—rowboat) and a semi-consonant or non-nuclear, lax rounded back vowel [ω] found after a vowel in an end-of-syllable position (e.g. لەۋ [læw]—lip). Given this double phonetic value, many people thought of using the letters *v* in the initial position and *w* in the end-of-syllable position. This, of course, would have gone against principle A as far as the correspondence between LSU, ASU and CSU was concerned, and it was decided to choose only one letter. The final choice was *w*, partly because it reads better in the end-of syllable position than *v*.

خ [x]: it was hard to find an ideal Latin letter to represent خ. Many people, used to PNSU, believed that the letter *h* would be the best choice and that it would also conform to the use of that letter in Chinese pinyin. Another argument in its favor was that it would avoid recourse to another compound letter, *kh*, commonly used in the West when transcribing Arabic- or Cyrillic-based languages. But *h* was used for another letter (ھ) and, in most languages, it is not used for the sound represented by خ. Some people insisted that, in the absence of a satisfactory solution, the letter *h* should be used for both sounds and the context would help discriminate[28]. Based on principle A, this was not acceptable. The

---

[26] In Chinese Pinyin *sh* denotes the voiceless palato-alveolar fricative(or domed post-alveolar fricative) [ʃ] and *x* denotes the voiceless alveolo-palatal fricative [ɕ] (or laminal postalveolar fricative). In Uyghur, there is only a voiceless palato-alveolar fricative [ʃ]. In ASU, both the PNSU *sh* and *x* are rendered through the letter ش.

[27] It is worth noting that, in PNSU, front and back vowels are not differentiated in a very systematic way: to the back vowels *a, o* and *u* correspond the front vowels *ə, ɵ* and *ü*: a new letter, a letter with an inner diacritical mark and a letter with an outer diacritical mark.

[28] The discussion sometimes took an amusing turn. An example is the discussion around the ambiguity of a

final choice was *x*, a letter commonly used in scholarly transcriptions and adopted by the Latin Azeri and Uzbek alphabets. This seemed to be the best option under the circumstances, though it could lead to potential double-spelling problems, with some foreigners finding, for example, the spelling *Xoten* (which English speakers may want to pronounce [zouten] and French-speakers [ksoten] or [gzoten]) less logical than *Khoten*. To this day, there are still Uyghurs who insist that this use of the letter *x* complicates things for XUAR Uyghurs who have to learn Chinese pinyin and therefore a second value for the same letter. For Uyghurs in the former Soviet Union (FSU), on the other hand, this letter is similar to the one used in CSU.

ژ [ʒ]: among the 11 propositions for this sound, none totally satisfied the requirements of all the rules. Here again, there was heated debate due to the fact that this sound is found in words of purely Turkic origin only as a spoken variant of ج [dʒ], in a pre-consonantal position (e.g. [taʒliq]—wearing a crown, from [tadʒ]—crown), and is therefore covered by the letter *j*. Independent use of this phoneme is limited to a few words of Russian or Persian origin. Even for these, it was felt that using the letter *j* would bring LSU visually closer to international use, since it is mostly found in words that are common to many languages (e.g. *jandarma*—French *gendarme*, *jurnal*—French/English *journal*, *injénér*—English *engineer*, *woltaj*—English *voltage*). Nevertheless, this was still in contradiction with principle A. Besides, it deprived LSU of an instrument to transcribe non-standard XUAR forms of the language, such as the Uyghur spoken in the FSU as well as the Ghulja and Qeshqer dialects. It was therefore decided to allow in yet another partially satisfactory double letter, *zh*, chosen for its use in Western transcriptions of Russian words. Familiar to English-speakers, this form is more difficult to read for speakers of other world languages. Besides being a double consonant, it can be confusing for XUAR Uyghurs since it is also a Chinese pinyin form used in parallel with *j* in PNSU for words of Chinese origin (e.g *fangzhen*—guideline). Supporters of *j* were not able to convince proponents of *zh* and, given that this phoneme occurs rarely in non-derivative context, it was decided to keep both tentatively and let usage decide.

ف	غ	ش	س	ژ	ز	ر	د	خ	چ	ج	ت	پ	ب	ئە	ئا
f	gh	sh	s	j/zh	z	r	d	x	ch	j	t	p	b	e	a
ي	ئى	ئې	ۇ	ئۈ	ئۆ	ئو	ھ	ن	م	ل	ڭ	گ	ك	ق	
y	i	é	w	ü	ö	u	h	n	m	l	ng	g	k	q	

Table 3. Arabic-Script Uyghur alphabet and its unified correspondence in Latin-Script Uyghur

The above table does not include the hamza on a tooth (ئ or ئـ), which was not discussed at the meetings. This letter is never listed separately in the ASU, due to the fact that it is considered an integral part of the initial form of vowels[29]. But, in fact, it does serve also

---

word spelled *haya* when using letter *h* indiscriminately: this word would transcribe both هايا [haja]— bashfulness and خايا [xaja], a rather coarse word to designate the penis.

[29] It is often said that the decision of Uyghur linguists to add this sign as part of the initial form of letters is a link with the old Uyghur writing system, in which all initial vowels were preceded by a tooth. The Arabic alphabet has 3 letters, ا, و and ي which can be used to indicate long vowels. Short vowels can be indicated through the use of vowel marks above or under the consonants but which are dispensed of in normal

an other purpose, that of indicating a glottal stop. In Uyghur, the glottal stop is not as strongly pronounced as it is in Arabic or in Uzbek, for example, and it has weakened to become no more than a hiatus. It appears usually in words of Arabic origin and replaces an original "ain" (ع) or a hamza (ء) in a median or final position (e.g. سائەت from Arabic وقوع [30]). The ساعَة, خائن from Arabic سۇئال from Arabic سُوَال, ۋۇقۇئ from Arabic خائن "hamza on a tooth" used in such context is traditionally not viewed as a separate letter but as a special spelling of vowels in median or final positions.

In most Latin transcriptions, the presence of a hiatus (and therefore of a hamza on top of a tooth inside a word) is indicated through the use of an apostrophe. The XSLC chose to follow the same tradition in its orthographic rules. At the same time, it was deemed unnecessary to do this in cases where there are two sequential vowels given that, according to Uyghur spelling rules (see below), only a *y* and a *w* can form a diphthong with the preceding vowel and, therefore, two sequential vowels can only belong to two different syllables (e.g. *qaide, aile, saet*). The hiatus-marking apostrophe is therefore reserved in LSU for use when a vowel is separated from a consonant by a hiatus (e.g. *qet'i, Qur'an*), when the two letters *n* and *g* have to be read separately and not [ŋ] as in *In'gliz*—English[31] (*baslan'ghuch, Hin'gan*), when *s* and *h* have to be read separately (*Is'haq*—Isaac), when *ng* and *h* have to be read separately (*Chong'haji*) and when the hiatus stands for a final   or   in literary words of Arabic origin (e.g. *toghra'*—royal seal, from Arabic   , *wuqu'*—phenomenon, from Arabic   ). This, of course, is not in keeping with the present spelling rules of ASU but it does not cause ambiguities and helps keep words shorter.

Basic LSU orthography and spelling rules

1. The first letter of a sentence and of a proper name is capitalized. For example : *Alim, Kanada, Uyghur, Ürümchi*.

2. The spelling of foreign proper names follows Uyghur spelling (as if transcribed directly from ASU) but the original spelling can also be added in brackets if necessary. For example: *Nyuyork* (New York), *Shenshi* (陕西/shǎn xī), *Skot Wéyd* (Scott Wade).

3. Abbreviations are capitalized, and separate letters can be followed by a point. For example: B.D.T (*Birleshken Döletler Teshkilati*, the United Nations), TDMP(*Teywen Démokratik Musteqqilliq Partiyisi*, the Taiwan Democratic Independence Party).

4. Only the first letter in a compound is capitalized. For example: AQSh (*Amérika*

---

writing. Given its phonetic characteristics, Uyghur notes down all vowels: ئا، ئە، ئو، ئۇ، ئۆ، ئۈ، ئې، ئى using derivates of traditional Arabic letters.

[30] The glottal stop in final position is usually not pronounced in the modern language and appears only in classical or religious texts .

[31] The existence of a single letter for [ŋ] in ASU, ڭ, differentiates clearly this sound from the group *n+g* written with two letters نگ.

Qoshma Shtatliri, the United States of America), Sh.U.A.R (*Shinjang Uyghur Aptonom Rayoni*, Xinjiang Uyghur Autonomous Region).

5. International units, assimilated foreign abbreviations and trade marks are written as in the original language, even if it means using letters that are not in the list of letters used for LSU. For example: *cm*, *kg*, *kv*, *VCD*, *MTV*, *Microsoft*, *Linux*.

6. As discussed above, the apostrophe is used to indicate a syllabic division when it is not otherwise obvious, that is when a vowel does not form a syllable with the preceding consonant. It is also used to differentiate the letter *ng* from the consonant group *n + g* and the letter *sh* from the consonant group *s + h*. Finally, it is used to indicate the etymological presence of a final glottal stop in some classical words. Unlike the *hamza* in ASU, it is not used between vowels to indicate syllabic division, since there can be only one vowel in a syllable. In this last instance, it is to be noted that in words directly transplanted from the phonetic groupings -*üen* (as in the Chinese unit of currency *Yüen* **يۇەن** or *shöyüen* **شۆيۈەن** , institute, and *gowuyüen* **گوۋۇيۈەن**, state council) and -*hua* (as in *Jungxua* **جۇڭخۇا**, Chinese) in modern Chinese ASU does not have an intervocalic *hamza*, thus turning -*ü*- and -*u*- into semi-consonants[32]. The LSU treatment of these words re-establishes actual Uyghur pronunciation, in which there is no initial semi-consonant.

7. Diacritical marks should be used whenever required since they help render sounds more exactly and prevent misunderstandings (*olturush*—to sit; meeting, party; *öltürüsh*— to kill; killing.

Present status and Future perspectives

The final version of the Latin-Script Uyghur alphabet was made public on July 3, 2001 (see table 3). Although LSU was originally developed exclusively for use in the computer field it has rapidly become widely accepted by educated Uyghurs and Uyghurologists for internet use as well as a transcription system for Uyghur in textbooks and other publications where it has basically replaced all other transcription systems.

More than five years of experimentation show wide acceptance among XUAR Uyghurs, Uyghur communities abroad and most Uyghurologists. Some Uyghur websites[33], to attract as wide a readership as possible, use all three transcriptions: ASU, CSU and LSU. Uyghur multiscript (ASU, CSU, LSU) converting tools[34] have rapidly developed to permit information sharing. As a result, people have started using LSU not only in emails or forums but also in web development, scientific research and application software development. People have come to realize that the script unification provided by LSU is not another imposed writing reform but an instrument which helps Uyghurs communicate and share information among themselves (XUAR Uyghurs, ex-USSR Uyghurs and

---

[32] Non-Chinese-speaking Uyghurs tend to pronounce these words *yen*, *shöyen*, *gowuyen* and *jungxuwa*.

[33] See biliwal.com, izdiyari.com, google.com/intl/ug/, rfa.org, oyghan.com, ukij.org, uyghurdictionary.org

[34] Online demo version is available at http://www.uyghurdictionary.org/tools.asp, offline plug-in version for Microsoft Word is available at http://oyghan.com/OTB/index.html . See also www.ukij.org for other tools.

Uyghur in other countries) as well as with other Turkic languages speakers around the world.

Apart from early inconsistencies due to entrenched habits, particularly in the shifts from PNSU *x* to LSU *sh* and PNSU *q* to LSU *ch,* a major problem in the use of LSU has been the reluctance of some computer users to use diacritics on the vowels *é, ö* and *ü*. This is due mostly to "keyboard laziness" as, in most cases, the absence of dedicated keyboards requires a double manoeuvre to input such letters (e.g., on the US international keyboard, diereses/accent + letter). This issue is being gradually solved due to the example given in major international sites using LSU.

Another problem is that usage has not yet totally solved the issue of *j* versus *zh* in the case of the ASU letter ژ , though most Uyghurologists have already opted for *zh*.

Conclusion

Regardless of the success of LSU, some people are still quite cautious, fearing it might eventually replace ASU. They argue that progress in the computer treatment of non-Latin writing systems, particularly with the advent of Unicode, has eliminated the need for a Latin transcription.

In the twentieth century, Uyghurs had more than their share of alphabet changes with the consequence that they have often been estranged from their cultural history, frustrated in their accumulation of knowledge and divided along political and generational lines by the artificial multiplication of writing systems.

As a clear and logical transcription system that is easy to learn and to use, LSU has boosted activities in Uyghur-language information processing and exchange and thus enabled Uyghurs to catch up with the world in their own language rather than in another language. Progress in ASU information processing and exchange can only benefit from the advances made through the use of LSU.

Further efforts are still necessary to develop proper instruments to deal with Uyghur-language information. Spelling mistakes due to the poor grasp of LSU as well as of ASU spelling rules are a significant problem for spell check and for multiscript converting tools. As low literacy rates and too many spelling reforms have caused a lack of consistent spelling among Uyghur speakers, refining inter-script conversion programmes and creating spell-check programmes have become priority tasks.

LSU has created a special niche for itself in the international community. A critical mass is forming around it. It has not replaced ASU but it is playing a supplementary role. It has enabled the Uyghur language to be present on the web and in publications in a readable way. Hopefully it will also gradually replace in the international press and on maps outdated transliterations or Chinese names for Uyghur personal and geographical name.

Acknowledgements

References

1. Alim Ahat, Frequency analyses of the Uyghur letters, Jan 2004.
http://www.uighursoft.com/uighur/soft/biliwal.htm

2. Waris A. Janbaz, *Kompyutér saheside "Latin yéziqi asasidiki Uyghur yéziqi" ni tesis qilish toghrisida teshebbusname* (29-page document distributed during the first conference), Nov 2000.

3. Uyghur Computer Science Association, "*Uyghur kompyutér yéziqi toghrisida*", http://www.ukij.org/teshwiq/UKY_Heqqide.htm , May 2004.

4. *UKY Arxipliri* (discussion archives on internet 200-2001)*,* www.biliwal.com/modules.php?name=UKY_Arxipi (also available at: www.voy.com/37018).

5. Website of the *XUAR Working Committee of Minorities' Language and Writing*, *wénzì gǎigé* (writing reforms), http://www.xjyw.gov.cn/han/wenzi_gai.htm, 2001.

6. J. R. Duval: "Modern Uyghur, A Historical Perspective." In: Culture Contact, History and Ethnicity in Inner Asia, № 2. Michael Gervers and Wayne Schlepp, eds. Toronto: Joint Center for Asia Pacific Studies, 1996, pp. 132-67.

Note: The online documents cited in this article may be subject to periodic hyperlink changes or loss. If this is the case, please contact the authors for archives.

Annex 1: Arabic-Script Uyghur, Cyrillic-Script Uyghur and Latin-Script Uyghur Alphabets

ف	غ	ش	س	ژ	ز	ر	د	خ	چ	ج	ت	پ	ب	ئه	ئا	ASU
f	gh	sh	s	j /zh	z	r	d	x	ch	j	t	p	b	e	a	LSU
ф	ғ	ш	с	ж	з	р	д	х	ч	ж	т	п	б	ә	a	CSU
ي	ئى	ئې	ۋ	ئۇ	ئۆ	ئو	ه	ن	م	ل	گ	گ	ك	ق	ASU
y	i	é	w	ü	ö	u	o	h	n	m	l	ng	g	k	q	LSU
й	и	е	в	ү	ө	у	o	h	н	м	л	ң	г	k	қ	CSU

Additional Cyrillic letters : ы ё ц э ю я