SCRIPT TRANSFORMATION SYSTEMS

Peeter Päll

Introduction

Standardization of geographical names is primarily done by establishing uniform spellings of names. It is easier and more efficient to standardize the graphic representation of names than e.g. the oral form, or pronunciation.

The existence of various writing systems would mean that there have to be standardized ways of representing geographical names originating in languages with other writing systems. It would be technically feasible to standardize names within a group of languages using the same writing system, e.g. those using Roman script. But for names from languages outside the group there are often different methods of representing them in Roman letters. One and the same non-Roman name might be rendered differently, depending on the target language, tradition and purpose of the text. A single Russian place name may serve as a good example of confusion created by various language-oriented renderings:

RUSSIAN	TRANSCRIPTION	LANGUAGE
Шахты	*Shakhty*	English
Šahty [UN]	*Šahtõ*	Estonian
ʃaxtɨ [phon.]	*Šahty*	Finnish
	Chakhty	French
	Schachty	German
	Sahti	Hungarian
	Szachty	Polish
	Sjachty	Swedish
	Şahtı	Turkish
	etc. etc.	

Although the name contains only five sounds, they have in this table 17 different graphic representations in Roman letters. The phoneme [ʃ] is represented by eight letters or letter combinations (*ch, s, š, ş, sch, sh, sj, sz*), [x] by three (*ch, h, kh*), and [ɨ] by four (*i, ı, õ, y*).

This is in sharp contrast with the use of names from the Roman-script languages where there are practically no variations in spelling for most of the names, exonyms exempted.

A solution that has been offered and is advocated by the United Nations is contained in the resolution of the first United Nations Conference on the Standardization of Geographical Names (1967, res. 9): one of the aims of international standardization is „to arrive at an agreement on a single romanization system, based on scientific principles, from each non-Roman alphabet or script, for international application". This is known as the **single romanization principle**. In other words, for each name written in a non-Roman alphabet there should be only one internationally accepted way of writing in Roman letters. In the case of the example above, this accepted spelling would be *Šahty*, according to a recommentation adopted in 1987.

Methods of conversion

Before going into other principles, it would be necessary to examine the technical terms involved. Definitions have been taken from the Glossary of Toponymic Terminology (see KADMON 2000: 299–326), when quoting resolutions of the United Nations the first part refers to the conference and the second part (after a slash) to the resolution number (see references in the end).

Conversion is a term denoting all changes that happen when phonological and/or morphological elements of a particular language are transferred to another language. Particularly this concerns transformations from one script into another. Names may be converted from Roman script into Arabic, from Thai into Russian, etc. Speaking of the direction of the conversion, there is the **source script** (also called donor script) – in terms of which a name is originally produced; the other is the **target script** (or receiver script) – into which a name is converted. **Romanization** is a conversion from non-Roman into Roman script, e.g. Greek Αθήνα → *Athína*, Russian Москва → *Moskva*, Arabic بيروت → *Bayrūt*, Hebrew תל־אביב → *Tel-Aviv*.

Conversion is effected by either transcription or transliteration. These two terms have earlier often been used without a clear distinction, partly this is true even today.

Transcription is a method of phonetic names conversion between different languages, in which the sounds of a source language are recorded in terms of a specific target language and its particular script, normally without recourse to additional signs called diacritics. Examples: Turkish *Ankara* ➞ Greek Αγκαρα, Russian Щукино ➞ English *Shchukino*.

Although transcription aims at giving the user a "pronounceable" name spelling, it is often only an approximation of the original pronunciation. Firstly, the sound structure of languages is different, and the sounds of a source language cannot always be recorded in the orthography of a target language. For example, an English speaker would not know how to pronounce certain Arabic sounds (ح, ص, ط, ع, غ) that do not have counterparts in English. Secondly, it is fairly usual that despite using an alphabet of a target language, an attempt is made to convey the written form of the source language, often without regard to pronunciation rules. A German name *Schmidt* is transcribed into Russian as Шмидт, not Шмит, although it is so pronounced.

This makes transcription linguistically imprecise and dependent on the orthography of a target language. Transcription is not normally a reversible process. If attempting to restore the source spellings of the examples given to the definition, the results could be different: Greek Αγκαρα ➞ Turkish *"Agkara"*, English *Shchukino* ➞ Russian "Шчукино".

The term "transcription" has a wider meaning also, signifying all kinds of notations. Phonetic transcription is a method of recording texts of a language as they are spoken, with the help of phonetic letters and symbols. Transcription can also be applied to so-called non-written languages. Logographic writing systems, such as Chinese, or Japanese Kanji, can only be transcribed phonetically into other scripts, since the source characters (logograms) represent morphemes (or words), not phonemes.

Although it is often thought that transcription is not used for conversions between Roman-script languages, there are some exceptions. Many of the geographical names that are nowadays viewed as exonyms have in fact been formed using the transcription method. Examples: Spanish *Zaragoza* ➞ English and German *Saragossa*, Italian *Toscana* ➞ German *Toskana*, Romanian *Bucovina* ➞ Polish *Bukowina*, French *Bruxelles* / Dutch

Brussel ➞ Hungarian *Brüsszel*, German *Hamburg* ➞ French *Hambourg*. There are also some languages where out of tradition and/or morphological considerations all names are transcribed. For example, in Latvian all nouns should have an ending in the nominative case, otherwise it would be difficult to decline the nouns. This has led to using Latvian orthography throughout the names, e.g. German *Stuttgart* ➞ *Štutgarte* (note the feminine ending *-e*), French *Tours* ➞ *Tūra*, English *Newcastle* ➞ *Ņūkāsla*.

Transliteration is a names conversion method between different alphabetic and/or syllabic scripts, in which each character (or character combination) is represented in the target script by one character (or character combination). Examples: Russian Челябинск ➞ *Čeljabinsk*, Amharic ትግራይ ➞ *Tigray*.

Transliteration, as distinct from transcription, aims at (but does not necessarily achieve) complete reversibility. **Reversibility** here means a characteristic which permits a written item to be converted from one script into another, and subsequently reconverted back into the source script, the result being identical with the original.

Transliteration key (also called conversion table) lists the characters of a particular source script together with the corresponding characters of a specific target script. In order to distinguish characters of the source script, ample use is made of all the character inventory of the target script, including, if necessary, **diacritics** (diacritical marks). These are small signs, placed above, below or across a letter or group of letters.

For example, Greek characters ι and η that are pronounced both [i] may be distinguished by adding a sub-macron to one of them: ι ➞ *i*, η ➞ *i̱*. Similarly distinction is made between Arabic ه ➞ *h*, and ح ➞ *ḥ*.

If full reversibility is aimed, in certain cases the transliteration alphabet becomes quite overloaded with diacritical marks. So in order to distinguish between five different consonants in Thai denoting *kh*, the ISO transliteration (11940:1998) makes use of several diacritics and their combinations:

ข ➞ k̄h, ฃ ➞ ḳh, ค ➞ kh, ฅ ➞ k'h, ฆ ➞ ḳh.

Transliterations of the International Organization for Standardization (ISO) are normally not used for rendering geographical names. They are mostly used in bibliography

and information processing when the representation of the exact original spelling is important. Sometimes these transliterations even disregard reading conventions. For example, in Thai certain combinations of a consonant and a vowel may be written in a

reversed order: the syllable *ke* is written เก in Thai and transliterated in the ISO system *ek*.

In contrast to ISO systems the ones recommended by the United Nations favour practical aspects and prefer systems that yield "pronounceable" names. As a comparison here are some names converted using a United Nations-recommended system and ISO transliteration:

	ORIGINAL	UN SYSTEM	ISO
Arabic	طَرَ ابُلُس	Ṭarābulus	ṭara'bulus
Bulgarian	Търново	Tărnovo	T''rnovo
Greek	Ναύπλιο	Náfplio	Naúplio
Hebrew	תֵּל־אָבִיב	Tel-Aviv	tėl-'åbiyb
Russian	Щельяюр	Ščel'jajur	Ŝel'âûr
Thai	ประเทศไทย	Prathet Thai	praethṣ̄ithy

In conclusion these two methods (transliteration and transcription) may be compared as follows:

	Transliteration	**Transcription**
Representing –	characters or syllabograms	sounds or phonemes
Information on –	original spelling	original pronunciation
Conversion rules are –	relatively easy	relatively complicated
Reproducing names and their pronunciation is –	often relatively complicated	relatively easy
Domains are –	scientific documentation	journalism, literature
Sources are –	philological works, library rules, ISO	orthographic dictionaries, pronunciation guides

(BACK 1997)

From national to international romanization

There are several steps that need to be taken, before an internationally adopted single romanization system is achieved.

Obviously, the first step would be to compile the romanization system to be applied. This is done on a national level, involving experts on linguistics, geography, etc. It is important that such systems are devised by experts in cooperation with those who would start to apply the system. Ideally the system would have to:
• be scientifically adequate, e.g. represent the phonological features or whatever unique properties of the language systematically,
• result in names that should be as easy as possible to write, read (pronounce), memorize and store electronically (includes minimal use of diacritical marks, avoiding difficult and unusual character sequences, etc.)
• be reversible,
• be simple and unambiguous to use (providing additional notes wherever there might be confusion),

Should some of the given characteristics be unattainable, it is the practical aspects of the system that matter most.

The second step is the national implementation of the system: using it in maps, on roadsigns, in official information, etc. The more widely the system is used, the better. It has been stated that new romanization systems for international use are considered only on condition that the sponsoring nations implement such systems on their cartographic products (maps and charts) (Res. IV/15).

After implementing the system nationally it is submitted for review to the United Nations Group of Experts on Geographical Names (UNGEGN). UNGEGN has a Working Group on Romanization Systems whose task is to study proposed systems. Sufficient time is needed for appropriate consultations and an expression of all views on technical matters between the sponsoring country and the Working Group. Specifically, the group considers whether the romanization system is based on sound scientific principles, the system's degree of reversibility, and the extent of its implementation on cartographical products by the proposing country.

When these discussions lead to a mutually satisfying conclusion, the romanization system is submitted in form of a resolution to the next full United Nations Conference on the Standardization of Geographical Names. When the resolution is adopted, the subject romanization system becomes the United Nations standard.

All is not over with the adoption of a resolution. UNGEGN and its working groups would have to monitor both national and international implementation of the system and if there are problems, report to UNGEGN sessions and the UNCSGN. Although it is highly desirable that states should refrain from revising systems previously adopted for international use (Res. IV/15), this sometimes does happen. The Working Group on Romanization Systems needs to be in contact with both the donor country and receiver countries to identify all problems of implementation.

It is also not a secret that not all of the adopted systems have been implemented, some of them even on a national level (e.g. the languages of the Indian group). From a point of view of receiver countries the implementation of some of the systems is complicated by high costs of switching over from the romanization systems that they used previously in mapping of foreign areas; names that were thus produced earlier, would need to be updated. For example, it was estimated that approximately 60% of the existing United States BGN Geographic Names Data Base file for Greece was outdated in terms of the ELOT 743 romanization system promulgated by the United Nations. (QUINTING 2000.)

The role of the United Nations: conclusion

Romanization is one of the important items on the agenda of the United Nations Conferences on the Standardization of Geographical Names. Since the very first conference in 1967 there has been a working group that addresses specifically the issues of romanization and prepares information for the United Nations Group of Experts on Geographical Names. At the beginning of 2002 resolutions had been adopted at the conferences that cover the romanization of 28 languages/scripts: Amharic, Arabic, Assamese, Bengali, Bulgarian, Chinese, Greek, Gujarati, Hebrew, Hindi, Kannada, Khmer, Macedonian Cyrillic, Malayalam, Marathi, Mongolian (in China), Nepali, Oriya, Persian, Punjabi, Russian, Serbo-Croatian Cyrillic, Tamil, Telugu, Thai, Tibetan, Uighur, Urdu. Other languages/scripts that are listed on the agenda include Armenian, Burmese, Byelorussian, Dzongkha, Georgian, Japanese, Kazakh, Kirghiz, Korean, Lao, Maldivian, Mongolian (Cyrillic), Pashto, Sinhalese, Tajik, Tigrinya, Ukrainian.

It must be noted that although romanization is the first priority in discussing script transformation systems, other types of name conversions are not excluded. UNCSGN resolutions have also addressed conversions from the Indian group of languages into Devanagari (II/11, III/12) and into non-Roman writing systems in general (particularly, Arabic, Res. IV/14). At sessions of UNGEGN and conferences information has been given also regarding name conversions into Hebrew and Russian.

A typology of conversions, including sub-types according to the availability of UN-recommended systems and the degree of implementation, is given in Annex.

Sources for further study

The UNGEGN **Working Group on Romanization Systems** website (http://www.eki.ee/wgrs/) contains information on non-Roman script languages and their romanization. The document *United Nations Romanization Systems for Geographical Names. Report on Their Current Status* has been updated to Version 2.1 (June 2002), due to be submitted to the Eighth UNCSGN in Berlin. The website contains also links to other sites devoted to romanization (ISO, ALA-LC, etc.).

The **UNGEGN** website (http://unstats.un.org/unsd/geoinfo/) provides background information on all activites of the Group, including various documents submitted to sessions of UNGEGN and conferences. The website contains links to UNGEGN working groups and national names authorities.

The publication by the United States Board on Geographic Names *Romanization Systems and Roman-script Spelling Conventions* (1994) lists some 29 BGN/PCGN romanization systems, including those languages that do not have any UN-recommended systems. Wherever needed, notes on the application of the systems are provided.

The most comprehensive publication on non-Roman (or, in fact, all) writing systems is that by DANIELS & BRIGHT 1996. Virtually all known scripts are covered with details on their inventory and the functioning of scripts, particular emphasis is on how the scripts render the actual phonology of the respective languages. Tables include transliteration equivalents and hints on pronunciation.

Requirements for romanization systems in detail are elaborated by QUINTING 2000. KADMON 2000 explains the issues concerning names conversion, transliteration and transcription.

References

Back, Otto. Fragen der Wiedergabe fremdsprachlicher geographischer Namen durch Exonyme oder durch Umschriftung. In: Kartographie und Namenstandardisierung. Wiener Schriften zur Geographie und Kartographie. Band 10. Institut für Geographie der Universität Wien, Ordinariat für Geographie und Kartographie, 1997, pp. 55–63.

(BGN.) Romanization Systems and Roman-script Spelling Conventions. Prepared by the U.S. Board on Geographic Names, Foreign Names Committee Staff. Published by the Defense Mapping Agency, 1994

Daniels, Peter T. & Bright, William (Ed.) The World's Writing Systems. Oxford University Press, New York–Oxford 1996. 922 pp.

Kadmon, Naftali. Toponymy. The Lore, Laws and Language of Geographical Names. Vantage Press, New York 2000. 333 pp.

Quinting, Gerd. Romanization Systems: Development and Application. In: Second International Symposium on Geographical Names GeoNames 2000. Frankfurt am Main, 28–30 March 2000. Compiled by Jörn Sievers with the assistance of Thaddäus Schneider. Edited by the Ständiger Ausschuss für geographische Namen (StAGN). Mitteilungen des Bundesamtes für Kartographie und Geodäsie. Band 19. Verlag des Bundesamtes für Kartographie und Geodäsie, Frankfurt am Main 2000, pp. 147–151.

Resolutions adopted at the seven United Nations Conferences on the Standardization of Geographical Names, 1967, 1972, 1977, 1982, 1987, 1992, 1998. Prepared for the United Nations by the Canadian Permanent Committee on Geographical Names. UNGEGN, 20th Session, New York, 2000. Information Paper No. 5

(WGRS.) United Nations Romanization Systems for Geographical Names. Report on Their Current Status. Compiled by the UNGEGN Working Group on Romanization Systems. Version 2.1. June 2002 [*presented to the Eighth UNCSGN, Berlin, 2002*]

Annex. **Conversion types and examples**

Type A. Non-Roman to Roman conversions

A1. UN-approved romanization systems

A1a: Nationally and internationally implemented

HEBREW	NATIONAL/UN
רִאשׁוֹן לְצִיּוֹן	Rishon-Leẕiyyon
רַמְלָה	Ramla
אַשְׁקְלוֹן	Ashqelon

A1b: Nationally implemented

BULGARIAN	NATIONAL/UN	BGN/PCGN
Хасково	Haskovo	Khaskovo
Ловеч	Loveč	Lovech
Велика Търново	Velika Tărnovo	Velika Tŭrnovo

A1c: Not implemented

MALAYALAM	ENGLISH
കോഴിക്കോട് Koḻikkoṭ	Calicut, Kozhikode
കൊച്ചി Kŏchchi	Cochin
തിരുവനന്തപുരം Tiruvanantapuraṁ	Trivandrum, Thiruvananthapuram

A2. No UN-approved romanization systems

A2a: National systems + systems in other countries

BYELORUSSIAN	NATIONAL	ENGLISH	ESTONIAN
Магілёў	Mahilioú	Mahilyow	Magiljov
Віцебск	Viciebsk	Vitsyebsk	Vitsebsk
Рэчыца	Rečyca	Rechytsa	Retšõtsa

LAOTIAN	NATIONAL + INTERNATIONAL	
ຈຳປາສັກ	Champasak (*ex*-Champassak)	
ສວັນນະເຂດ	Savannakhèt	
ໄຊຍະບຸລີ	Xaignabouli (*ex*-Sayaboury)	

A2b: National system, no international use

DZONGKHA	NATIONAL	CONVENTIONAL
བུམ་ཐང	B'umtha	Bumthang
གཞལ་སྒང	Zh'ämgang	Shemgang

A2c: No national system, various systems in other countries

ARMENIAN	ENGLISH (BGN/PCGN)	ESTONIAN
Երևան	Yerevan	Jerevan
Չարենցավան	Ch'arents'avan	Tšharentshavan
Վայոց Ձոր	Vayots' Dzor	Vajotsh Dzor

KAZAKH	ENGLISH (BGN/PCGN)	ESTONIAN
Қостанай	Qostanay	Kostanaj
Маңғыстау	Mangghystaū	Manggõstau
Сарыарқа	Saryarqa	Sarõarka

A2d: No national or known international systems

SINHALESE	TAMIL	ENGLISH
කොළඹ Kŏḷamba	கொழும்பு Kŏḻumpu	Colombo
යාපනය Yāpanaya	யாழ்ப்பாணம் Ȳāḻppāṇam	Jaffna
මහනුවර Mahanuvara		Kandy

Type B. Roman to Roman conversions

ESTONIAN	LATVIAN
Aardla	Ārdla
Vääna	Vēna
Püssi	Pissi

Type C. Roman to non-Roman conversions

ESTONIAN	RUSSIAN
Aardla	Аардла
Vääna	Вяэна ("Vjaèna")
Püssi	Пюсси ("Pjussi")

Type D. Non-Roman to non-Roman conversions

UKRAINIAN	RUSSIAN
Хмельницький (Khmel'nyts'kyi)	Хмельницкий (Hmel'nickij)
Суми (Sumy)	Сумы (Sumy)
Мелітополь (Melitopol')	Мелитополь (Melitopol')