



Script transformation systems

Peeter Päll

Wien, 17.03.2006

Single romanization principle

- Names originating in Roman-script languages are usually unchanged (e.g., *Västerås*)
- Names originating in non-Roman-script languages are **converted** into Roman script which can be done differently, e.g. Russian Шахты (Šahty):
 - ш – ch / s / š / ș / sch / sh / sj / sz
 - х – ch / h / kh
 - ы – i / ı / õ / y
- One non-Roman name – one Roman spelling

Conversion

- Source script (donor script) → Target script (receiver script)
- **Romanization** – conversion into Roman script

Transcription

- *Phonetic conversion* between different languages, conveys sounds in terms of a target language and its script
 - sounds or phonemes
 - “pronounceable” spelling
 - (often) complicated rules
 - non-reversible
- Other meanings:
 - phonetic transcription
 - notation of speech

Transliteration

- *Graphic conversion* between scripts, conveys characters in terms of a target script
 - characters, syllabograms
 - “difficult” spellings (incl. diacritical marks)
 - often simple conversion rules
 - (ideally) reversible

Reversibility

- Russian ц [ts] → с (not “ts”) → ц (not “тс”)
- Ельцин → El’cin → Ельцин
- cf. Ельцин → Yel’tsin → Йельтсин?
- Арбатская → Arbatskaја → Арбатская

Diacritical marks

- acute á, é, í, ú
- cedilla ç, ç, ñ
- circumflex â, ê, î, û
- diaeresis ä, ï, ü
- grave à, è, ì, ù
- macron ā, ē, ī, ō
- tilde ã, ñ, õ

■ Using diacritical marks for distinction:

- Greek ι → i, η → ï

From national to international romanization

- 1. Preparation of the (national) romanization system:
 - scientifically adequate
 - easy to write, pronounce, memorize, store
 - reversible
 - easy to apply
- 2. National implementation of the system
 - using in maps, documents, on road signs, etc.; sufficiency of names
- 3. Submission to UNGEGN for consideration
 - Working Group on Romanization Systems
- 4. Resolution at the UN Conference on the Standardization of Geographical Names
 - system becomes the UN standard
- 5. Monitoring of the implementation

United Nations romanization systems for geographical names

■ **Nationally and internationally implemented:**

- Amharic*, Arabic*, Chinese, Hebrew, Khmer*, Macedonian Cyrillic, Mongolian (in China), Persian, Serbian, Thai, Tibetan, Uighur

■ **Nationally implemented:**

- Bulgarian, Greek, Russian

■ **Not implemented:**

- Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Nepali, Oriya, Punjabi, Tamil, Telugu, Urdu

■ *No UN system, some national systems in use:*

- Byelorussian, Dzongkha, Georgian, Japanese, Korean, Maldivian, Ukrainian

■ *No UN or national system:*

- Armenian, Burmese, Kazakh, Kirghiz, Lao, Mongolian (Cyrillic), Pashto, Sinhalese, Tajik, Tigrinya

Sources for further study

- Working Group on Romanization Systems
 - <http://www.eki.ee/wgrs/>
 - *UN Romanization Systems for Geographical Names* (2003)
- UNGEGN website
 - <http://unstats.un.org/unsd/geoinfo/>
- US BGN / PCGN systems
 - *Romanization Systems and Roman-script Spelling Conventions* (1994)
- Daniels, P. & Bright, W. (Ed.)
 - *The World's Writing Systems* (1996)