

Introduction. Reversibility statistics

WGRS meeting
18 May 2012
Gdańsk

Resolution I/9

- 1. Recommends that efforts should be made to arrive at an **agreement on a single romanization system**, based on scientific principles, from each non-Roman alphabet or script, for international application;
- 2. Further recommends that a **special working group** under the responsibility of the proposed United Nations Permanent Committee of Experts on Geographical Names should concentrate on this subject ...
- 3. ...

Resolution IV/15

- 1. Recommends that new romanization systems for international use should be considered only on condition that the sponsoring nations **implement** such systems on their cartographic products (maps and charts);
- 2. Further recommends that States should **refrain from revising** systems previously adopted for international use.

Resolution IX/8

- Recognizing, however that, subsequent to the international adoption of a romanization system, the sponsoring nation or nations may, for reasons unforeseen at the time of adoption, experience difficulty in achieving national implementation of that system,
- Recommends that the United Nations Group of Experts on Geographical Names be able to **reconsider** the relevance of any internationally adopted romanization system that has not been officially implemented by the sponsoring nation or nations within **ten years** following the adoption of the relevant resolution, or that is no longer implemented by that nation or those nations after that same period of time.

Reversibility (of romanization)

- = the ability to reconstruct on the basis of a romanized form unambiguously the original-script form
- Reversibility is not achievable if the character repertoire of the source script is significantly larger than that of the Roman script
 - e.g. logographic script

not reversible

- Chinese, Mongolian (in China), Thai, Tibetan
 - e.g. in Thai script there are 5 characters denoting the sound [kh]
 - romanization of Thai does not indicate vowel length or intonation

(not) reversible

- Greek (transcription variant)
 - no distinction between ι and η (=i), ο and ω (=o)

..generally not reversible..

- Khmer
 - the script is complex, and does not indicate word division
 - one and the same sound may be represented by different characters in the Khmer script

..not fully reversible..

- Hebrew, Urdu
 - in Hebrew ו and נ are romanized **t**, ו and ף are romanized **k**; vowels (e.g. **o**) have more than one way of writing
 - in Urdu some of the Arabic consonants (e.g., ح and ة) are undifferentiated in romanization

..mainly reversible..

- Amharic
 - some syllables of the 1st order are pronounced in the same manner as the syllables of the 4th order
 - some consonants are undifferentiated

..mostly reversible..

- Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, (Marathi), (Nepali), Oriya, Punjabi, Telugu
 - there *may* exist some ambiguities in the romanization of vowels (independent vs. abbreviated characters) and consonants (ligatures vs. character sequences)

..generally reversible..

- Arabic, Persian
 - there are some ambiguous letter sequences (*dh*, *kh*, *sh*, *th*) which may also point to combinations of Arabic characters in addition to the respective single characters
 - there are some ambiguous letter sequences (*kh*, *sh*, *th*, *zh*) which also may represent the romanized values of two Persian characters in addition to the respective single ones

..on the whole reversible..

- Bulgarian, Macedonian Cyrillic, Tamil
 - Bulgarian: one should know the spelling rules. For example, the romanized *j* will correspond to *ь* if used after a consonant, in other cases it represents the Cyrillic *й*. As exception, when *j* is followed by *a* or *u*, the combinations *ja* and *ju* should be converted to the Cyrillic as *я* and *ю*, respectively. These romanizations could be ambiguous, if there should exist character sequences *йа* and *йу* which seems highly unlikely
 - Mac. Cyr: occasionally romanized characters *dz*, *dž*, *lj* may represent character combinations in Cyrillic (дз, дж, лј) instead of single characters (с, ц, љ). Also, both characters *r* and *к* or *ř* and *к* may occur before *e* and *и*

..reversible, but..

- Russian
 - very rarely there can be ambiguities. For example, a geographical name *Kazenojam* of Daghestan is reconstituted as *Казенойам* in the Cyrillic source script, not *Казеноям* as would seem logical
 - (*ja* = *я*, rarely *йа*; *ju* = *ю*, rarely *йу*)

..unambiguous, but..

- Uighur
 - in the form intended for general use does not differentiate some of the Uighur characters (**e, g, h, k, o**)

reversible

- Greek (transliteration variant), Serbian

Summary of UN systems

- Of 28 adopted systems:
 - 4 are not reversible (bo, mn-CN, th, zh)
 - 1 is (not) reversible (el-transcr)
 - 1 is generally not reversible (km)
 - 2 are not fully reversible (he, ur)
 - altogether 8 systems (29 %)
 - 17 are mainly/mostly/generally/on the whole reversible (am, ar, as, bg, bn, fa, gu, hi, kn, mk, ml, mr, ne, or, pa, ta, te)
 - 2 are reversible (unambiguous), but.. (ru, ug)
 - 1 is reversible (sr)
 - altogether 20 systems (71 %)

Proposed systems

- Arabic (2007)
 - generally reversible
- Bulgarian (2006)
 - no distinction between *a* and *ъ* (=а), perhaps some ambiguous character sequences
- Byelorussian/Belarusian (2007)
 - generally reversible though there may exist some ambiguous consonant combinations
- Persian (2003/2012)
 - not fully reversible (cf. Urdu)
- Ukrainian (2010)
 - certain character combinations may cause ambiguity and some characters are not differentiated in romanization

Which criterion is more important?

- Reversibility
 - scientific approach
- Practical implementation in donor country
- Practicality (simplicity) of the system
 - use of diacritical marks
- “Language-neutrality” of the system
 - English vs. other (Slavic, etc.) languages