**United Nations Group of Experts on
Geographical Names
Second session**
New York, 3 – 7 May 2021
Item 15 of the provisional agenda [*]
Writing systems and pronunciation

Transcriptor: Solving your transliteration problems?

Submitted by the Netherlands[**]

The UNGEGN propagates the use of local spelling practices in the international exchange of geographical descriptions. This clashes with the current use of exonyms: some languages have many, others have few, but they exist in virtually every language, where they are felt as common and familiar.

Due to globalization, people are confronted with many more geographical names than a few decades ago. Moreover, due to new political developments, new names appear in media, others are forgotten. The consequence is that people such as news reporters or business (wo)men, have a need for proper rendering of place names from foreign countries, in particular of names for which neither an established transcription nor exonym exist. The local names policy is a simple heuristic in case of corresponding spelling systems and alphabets, even when in such instances unusual diacritics may cause difficulties in easy writing/typing, not to mention pronunciation. Things get even more complicated when different spelling systems are involved. From the Dutch perspective, Cyrillic, Greek or Arab writing systems require some form of transliteration or transcription.

Here the difficulties for the common user are manifold:

- The transliteration keys are not always stable over time;
- Countries with the same or very similar alphabets use different transliteration keys;
- Some keys contain many diacritics, alien to Dutch (not to mention English) users;
- The result may be a concatenation of characters that will not lead to something even close to a proper pronunciation of the name.

Depending on the purpose of the text, the user may have various needs: an internationally and locally accurate form for diplomatic use or rather travel purposes, an easily processable form or even an exonym for use in the context primarily addressing people being member of the speech community of the author.

The *Transcriptor* aims to provide information about the spelling of names, originally written in a non-Latin alphabet. It is a web application consisting basically of four different modules:

- A transliteration algorithm, that can apply the current official transliteration rules per alphabet and country;
- A transcription algorithm that can convert the official transliteration into a spelling, closer to the actual spelling conventions of the receiving language, in casu Dutch, avoiding diacritics or digraphs that will lead to severe mispronunciations;
- A dynamic database module, searching various online sources, such as Wikipedia or GeoNames, for existing spelling practices in various languages;
- A list of current Dutch exonyms.

An important asset of the tool is that it provides the user with various options, from which a choice can be made that fits the needs of the audience of the writing.

In order to be able to find a proper form, one needs a search string as a start. The available source information can consist of various media: a map, piece of writing or picture or a shield, showing the name in the original alphabet form. Such a text string can be copied and pasted into the *Transcriptor*. But often, only a version from a foreign news agency, a local business partner who came up with an own transcription or the like, is the source. The *Transcriptor* can search two ways: from the original to a transliteration/transcription or from a text string to the original. For the latter operation, it needs some hints from the user (which country, is it a place name or a proper name, etc.) and uses the databases it has access to in order to identify the object or person and the corresponding original spelling, which is supposed to be the basis for a subsequent proper transliteration or transcription.

The first release of the *Transcriptor* covered the Cyrillic alphabet in Russia and Ukraine, with important differences. The next version is built for Arabic, which has many complications for Dutch users, due to the many different countries where Arabic is an official language, sometimes with different

transliterations keys, the regularly lacking vowel characters or the various existing spelling traditions in Latin alphabets for major (historical) places or people, which for some countries is based on English, for others on French spelling practices. Because of the unfamiliarity of most speakers of Dutch with Arabic pronunciation, a user-friendly transcription is one of the desiderata.

The *Transcriptor* can be used both for geographical names and proper names. Future extensions are the inclusion of Greek, Hebrew and furthermore Southeast Asian writing systems. Much of its engine is target language independent, which may offer the opportunity to apply this system in future for other languages.

The *Transcriptor* is developed by a team of IT and language specialists, led by Nicoline van der Sijs (*Instituut voor de Nederlandse Taal* or Dutch Language Institute, Leiden). The implementation of the Arabic module is sponsored by the *Nederlandse Taalunie* (Dutch Language Union). The *Taalunie* is responsible for the Dutch spelling and intends to offer the *Transcriptor* on their webpage, alongside information about established Dutch exonyms, which can be considered part of the Dutch vocabulary, requiring the application of Dutch spelling principles.

Points for discussion

The Group of Experts is invited to:

1. Take note of the development of the Transcriptor web application;
2. Comment and provide input on its principles and characteristics as described in this report.