



ISSN: 2350-0328

**International Journal of Advanced Research in Science,
Engineering and Technology**

Vol. 8, Issue 10 , October 2021

Unstable combination of transliteration and transcription in the lexicon and phonetics of the Uzbek language

Norov A.M., Saidova D.E., Jorabekov T.K

PhD, Department of Algorithms and Programming Technologies, Karshi State University, Karshi, Uzbekistan
Applicant, Department of Algorithms and Programming Technologies, Karshi State University, Karshi, Uzbekistan
Applicant, Department of Applied Mathematics, Karakalpak State University, Nukus, Uzbekistan

ABSTRACT: In Uzbek linguistics, certain rules are introduced for transliterating words in the Cyril-Latin and Latin-Cyrillic directions and these rules are almost not difficult to construct a transliteration algorithm. However, there are some exceptions related to the phonetic word and the transcription of this word, which, in turn, require specific complex methods in the construction of the transliteration algorithm. Therefore, in this article, we have considered the issue of computer modeling of transliteration for phonetic words.

KEY WORDS: Uzbek Lexicon, Lexical parts, Transliteration, Retransliteration, Transcription, Uzbek Phonetics, Orthotranslit.

I. INTRODUCTION

Due to the publication of the Law of the Republic of Uzbekistan “On the introduction of the Uzbek alphabet based on the Latin script” and the Resolution of the Cabinet of Ministers of the Republic of Uzbekistan “On approval of basic spelling rules of the Uzbek language”, Cyrillic and Latin scripts are still used in parallel. Spelling rules based on Cyrillic graphics have been in force in our country since 1956. The spelling rules in Latin script were adopted in 1995, and these two spelling rules differ significantly from each other.

The spelling of words with such formal changes in both graphics creates specific difficulties in mastering the grammar of the Uzbek language, typography related to text editing, transliteration and, in general, the effective use of information technology in some areas of Uzbek linguistics. Because this situation requires, on the one hand, spelling literacy in language, on the other hand, special literacy in information technology.

We clarify our views on this by describing the problems that arise in the process of direct (Cyrillic-Latin) and inverse (Latin-Cyrillic) transliteration.

II. SIGNIFICANCE OF THE SYSTEM

The paper mainly focuses on how to use the transcription rules when transliterating and retransliterating the Uzbek text in the direction of Cyril-Latin and Latin-Cyril. The study of literature survey is presented in section III, Methodology is explained in section IV, section V covers the experimental results of the study, and section VI discusses the future study and Conclusion.

III. LITERATURE SURVEY

Transliteration can be seen as a model that plays a key role in translating people’s names and literature names into another alphabet, creating materials for computer-assisted language teaching, transcribing foreign words, and so on. Many scientists have done research on this.

In their book, Kunchukuttan and Bhattacharya talk about improving machine transliteration using script similarities between multilingual and related languages, and proposed a multilingual neutral transliteration model for transliteration that includes relevant languages.



ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 10 , October 2021

LevShcherba analyzed the phonetic features of the Russian letter “ц” in English, Russian, German and French. According to him, the Russian sounds “ц” (“c” in English, “z” in German) and “ts” in are different. As an example, he shows how a syllable is separated by the following words: отсыпке (at/syp/ke) and оцпке (a/syp/ke). L.Shcherba also notes that in his work there is no exact equivalent of the consonants “ц”, “ч”, “ш”, “ж”, “щ” and “х” in the Russian alphabet in the Latin alphabet.

VilenKomissarov interprets the concepts of transliteration and transcription as mutually exclusive and complementary. He argues that transliteration and transcription is a way of translating lexical units while preserving their original character. In transcription, the sound (pronunciation) form of a word is formed, while in transliteration, its graphic form is displayed. Therefore, transcription, which retains some elements of transliteration, is one of the leading methods in modern translation practice.

IV. METHODOLOGY

It is well known that transliteration is the literal representation of a text or individual words written on the basis of one alphabet by means of another alphabet.

However, if we take into account the words that make up the existing modern vocabulary of the Uzbek language and the phonetic features of these words, the features of transliteration described above are not enough for our present study. Since the Uzbek dictionary literally consists of three parts of vocabulary (its own lexical part, a part assimilated from Eastern languages and a part assimilated from the Russian language), each word in these parts has a special place in transliteration.

Indeed, in some sources we find other definitions of transliteration and transcription. For example, they are given below:

Transliteration is the process which consists of representing the characters of an alphabetical or syllabic writing by the characters of a conversion alphabet.

Retransliteration is the process whereby the characters of a conversion alphabet are transformed back into those of the converted writing system. It is the exact opposite of the transliteration process in that the rules of a transliteration system are applied in reverse in order to reconvert the transliterated word to its original form.

Transcription is the process whereby the pronunciation of a given language is noted by the system of signs of a conversion language.

The 80,000-word explanatory dictionary of the Uzbek language contains about 600 words with the Cyrillic letter “ц”. In addition to this dictionary, the number of words associated with this letter in our lexicon will increase even more if we take into account people’s names, geographical place names, and some new scientific terms.

Thus, non-uniform operations are common in both Cyrillic-Latin transliteration and Latin-Cyrillic retransliteration. Therefore, we need a special transliteration that combines the Uzbek spelling rules in Cyrillic and Latin script as fully as possible. For such transliteration, we have introduced the concept of “Orthographic transliteration” or “Orthotranslit” for short.

Orthotranslite is an expanded form of transliteration that incorporates the principles of spelling, orthoepy, phonetics, morphology, and morphonology.

In other words, the main difference between Orthotranslit and simple transliteration is that letters or letter combinations are not simply transferred from one alphabet to another, but are made based on spelling rules adopted for a particular natural language.

Here is the text of the algorithm for expressing words in the Cyrillic alphabet in the Latin alphabet using orthotranslite. This algorithm consists of the following steps:



ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 8, Issue 10 , October 2021

Step 1: An array of independent words (ArrP) with the Cyrillic letters ‘y’ and ‘ю’ at the end is created and they are temporarily excluded from transliteration.

Step 2: The “u” and “yu” particles in any word structure other than the ArrP array are separated from the words.

Step 3: Create an array of uppercase letters (CyrCptV) and an array of lowercase letters (CyrSmlV) in the Cyrillic alphabet.

Step 4: Create an array of capital consonants in the Cyrillic alphabet (CyrCptC) and an array of lowercase consonants (CyrSmlC).

Step 5: The graphemes “E”, “Ё”, “Ю” and “Я” that appear in initials and abbreviations in people’s names are all transliterated to the letter “Y” in the same order.

Step 6: Transliteration of chronological dates given in Arabic and Roman numerals in the text is performed.

Step 7: General transliteration is performed for all remaining cases.

An algorithm for transliterating Uzbek words given in the Latin alphabet into words in the Cyrillic alphabet (retransliteration) based on spelling rules will also be developed.

V. EXPERIMENTAL RESULTS

Of the 584 words in the Uzbek lexicon, 444 can be transliterated without reference to any database, but the remaining 140 words are almost completely transliterated (Table 1).

Table 1 Fragments with the Cyrillic letter “ц”

Frag.	Quant.	Frag.	Quant.	Frag.	Quant.	Frag.	Quant.	Frag.	Quant.
кция	34	кцио	17	порци	4	цент	10	цикл	10
нция	25	цио	47	нциа	3	оце	9	цито	6
ция	252	крипци	2	циа	5	це	15	цид	5
Total:	311	Total:	66	Total:	12	Total:	34	Total:	21

In many words from the Russian language, letters other than “ц” are also actively involved, for example “Ъ”, “ь”, “ля”, “лю” and etc. In the process of transliteration from Latin to Cyrillic, complex algorithms have to be developed to preserve the original form of such words.

VI. CONCLUSION AND FUTURE WORK

Development of a model of transliteration and transcription for the Uzbek language writing in this alphabet some concepts and terms that belong to the peoples of the world and are internationally recognized, has a positive effect on their phonetic convergence and correct pronunciation.

REFERENCES

- [1] Kunchukuttan, A. and Bhattacharyya, P. (2021). Machine Translation and Transliteration Involving Related and Low-resource Languages. CRC Press is an imprint of Taylor & Francis Group, LLC.
- [2] Komissarov, V.N. Translation theory (linguistic aspects). Moscow, Higher school, 1990. 253 p.
- [3] Shcherba, L.V. Selected works on linguistics and phonetics. Vol. 1. Moscow, 1958. 184 p.
- [4] Norov, A. (2020). Computer-oriented Models of Uzbek Linguistics. PhD thesis, Karshi State University.