A NEW APPROACH TO AUTOMATED AZERBAIJANI-ENGLISH ATRANSLITERATION

S. Mammadzada, Institute of Information Technology of ANAS, Baku, Azerbaijan, sabina_ict_az@mail.ru

The paper highlights the importance and dominance of the English language in the globalizing world. English is explored as a lingua franca. Moreover, the necessity of a new approach to Azerbaijani-English transliteration is emphasized. In this regard, new conversion table for Azerbaijani-English language pair is set out. The transliteration principles and transformation rules for automated transliteration system are defined and the accuracy of this system is provided. The possible contribution of the proposed transliteration system for machine translation system of Azerbaijani-English language pair is shown.

Key words: *transliteration system; Romanization; conversion table; machine translation.*

НОВЫЙ ПОДХОД К АВТОМАТИЗИРОВАННОЙ АЗЕРБАЙДЖАНО-АНГЛИЙСКОЙ ТРАНСЛИТЕРАЦИИ

С. Маммадзадэ, ***Институт информационных технологий НАНА, Баку, Азербайджан,*** **sabina_ict_az@mail.ru**

В статье подчеркивается важность и доминирование английского языка в глобализирующемся мире. Английский изучается как лингва франка. Кроме того, подчеркивается необходимость нового подхода к азербайджано-английской транслитерации. В связи с этим намечена новая конверсионная таблица для азербайджано-английской языковой пары. Определены принципы транслитерации и правила трансформации для автоматизированной системы транслитерации и обеспечена точность этой системы. Показан возможный вклад предложенной системы транслитерации в систему машинного перевода азербайджано-английской языковой пары.

Ключевые слова: *система транслитерации; латинизация; таблица пересчета; машинный перевод.*

1. Introduction

Textual exchanges are performed through mail, instant messaging, real-time chat protocols, asynchronous discussion forums, Web pages, and etc. all of which constitute Computer-Mediated Conversation. In the global network, this ‘conversation’ is mainly realized in English. This fact has several reasons, one of

which is the emergence of the recent term ‘English as a lingua franca’ (ELF). This is a way of regarding the communication in English among the speakers, whose first language is non-English. According to Crystal’s estimations, approximately only one out of every four users of English in the world is a native speaker of the language (Crystal, 2003). It means that the greatest part of ELF communications occur among ‘non-native’ speakers of English. English, in most cases is ‘a ‘contact language’ between persons who share neither a common native tongue nor a common (national) culture, and for whom English is the chosen foreign language of communication’ (Firth, 1996).

Most linguists indicate that English is the most broadly used language for international and intercultural communication. The report presented by Ammon on the global usage of English illustrates the current prevailing position of English. He states that the dominance of English s confirmed by the statistics, (1) English is spoken by approximately 1,5 billion people worldwide, (2) English is adopted as an official language of 62 nations, (3) about 70-80 percent of academic publications in scientific communication is written in English, (4) in most international organizations, English is the de facto official and working language, and (5) English is the most taught foreign language throughout the world (Ammon, 1992).

At present, non-English-speaking users are eager to search information on the web, particularly, researchers are very enthusiastic to explore and publish articles to English-language journals. They try to access to a broader audiences and scientific academic community. Sunol and Saturno believe that the studies conducted in languages other than English are not accessible and cited as much as those written in English (Sunol, Saturno, 2008).

The notion of transliteration is explained as a process of replacing or supplementing the scripts or words of one language with the scripts or words of another. The key point here is the exact or at least the maximum equivalence of the phonology of the source language in the target language. Transliteration helps to establish the context in which cross-cultural translation is recognized. Nida and Taber state that untranslatable words usually appear when exact equivalence of the word does not exist in target language and comparative equivalence is needed (Nida, Taber, 1969). In this case, transliteration is used.

Obviously, most web resources are Romanized for being accessible for a wider audience. Numerous Romanization tables are adopted throughout the world in this regard. Although, several conversion tables for the Azerbaijani language are adopted by international and local organizations, the proper international standard for this language has not been accepted yet (Library and Archives Canada, Ottawa, 2006). The first of the three standard adopted by ISO9 is on the transliteration of Arabic characters (ISO 233: 1984), which were used in Azerbaijan at that period (1984), into Latin characters. The second standard was adopted in 1993 as the second edition of ISO 233 to simplify Arabic

transliteration; however these scripts had already been disabled in Azerbaijan prior to 1993. The third and the foremost transliteration standard ISO 9 was adopted in 1995 for the transliteration of Cyrillic characters into Latin characters (for Slavic and non-Slavic languages).

Since the problem of transliteration may have an effect of the research activity of many students and researchers, its study and development of methods becomes significant. Many researchers have proposed and developed several transliteration methods for separate languages so far. Though a few transliteration tables for Azerbaijani language are adopted by some organizations, their modification and standardization problems still remain. The modification of the transliteration tables comprises the exclusion of diacritics from the table and replacement of some letters, the phonology of which does not appropriate for the phonology of Azerbaijani language. This study attempts to fill the gaps existing in Azerbaijani-English transliteration.

2. Azerbaijani-English transliteration challenges and their solution

The alphabet of Azerbaijani language includes 32 Latin letters, 7 (ç, ğ, ı (lower case for «I»), İ (upper case for «i»), ö, ş, ü) of which are modified and 1 letter (ə) is adopted specifically for the Latin alphabet of Azerbaijani language. In fact, these letters become very challenging in the process of transliteration, since not only their graphical description, but also their pronunciations are unknown for English-speaking audience. Representation of these letters by the diacritical signs in adopted transliteration tables, such as ġ, ž, ǰ, ̌, ï, ̂, ù, š, č, ĉ, which are unfamiliar even for Azerbaijani reader, makes this problem even more complicated. The representation of the Azerbaijani letter «ə» is another problem. This letter is adopted in few languages, such as Abkhaz, Kazakh, Tatar, and Bashkir, although pronounced differently or not pronounced at all. Although, several attempts have been made for the unification of all Turkic alphabets during the past two decades in order to ease the writing system of all Turkic languages using Latin alphabet, adoption of common Turkic alphabets has failed (Bedii Duru Altuğ, 2014).

The problem of lack of letters also exists in the transliteration from English into Azerbaijani. For example, the letter «w» does not exist in Azerbaijani alphabet and has to be replaced by letter combination «ou» or «v».

One of the problems complicating transliteration issue of Azerbaijani-English language pair is the significant difference in the number of the scripts in both alphabets. Since the number of letters in the English alphabet is less than in the Latin alphabet used for Azerbaijani language, the use of letter combinations for the transliteration is required in order to accomplish a comprehensive transliteration (Weinberg, 1974).

Azerbaijani		English	Azerbaijani	English
A		A	Q	G
B		B	L	L
C		J	m	m
Ç		ch	n	n
D		D	o	o
E		e/a	ö	o
ə		E	p	p
F		F	r	r
G		G	s	s
Ğ		gh	ş	sh
H		H	t	t
X		kh	u	u
I		I	ü	u
I		I	v	v
J		Zh	y	y
K		K	z	z

The use of letter combination is another challenging issue. Thus, letter combinations adopted for the letter with the same sound vary depending on the language. For example, the Cyrillic letter «я» can be transliterated as «ia», «ja», and «ya» depending the conversion system. The point is that the characteristics of specific phonemes in Slavic languages are not written explicitly with only one Roman letter. Therefore, several equivalence for Azerbaijani letter «y» arises, which include «i», «j», and «y».

Other problem related to the transliteration is adoption of the conversion tables for the ancient alphabets, which are not used anymore. Adoption of these transliteration schedules is significant for the libraries, where numerous literary materials written in ancient alphabets are preserved. Although they have not been used for many decades, or even centuries, the phonological and graphic features of these letters are still taken into account in the design of the new conversion schedules. This fact complicates the modification, simplification, and finally, the improvement of available transliteration standards.

Transliteration problems also arise due to some political and social factors. For example, the territory of modern Azerbaijan has been invaded by different governments several times. Each invasion has brought new culture, traditions, languages and scripts with itself. With the conquest of Arabic Khalifat and the spread of Islam in Azerbaijan gave impetus to the rise of Arabic script here and the ancient Latin script used here was replaced by the Arabic one. Whereas in the period of Soviet Union, Arabic script was replaced by the Cyrillic. In addition, in different periods, the mentioned Cyrillic alphabet has been modified for several times to achieve the correspondence with the phonological features of Azerbaijani language. After gaining independence in 1991, Azerbaijani government adopted modified Latin script for Azerbaijani language (Law of the Republic of Azerbaijan on the renewal of the Azerbaijani alphabet with Latin graphic, 1991). The key point is that, in each period, numerous literary manuscripts, political documents and materials have been generated in each of these scripts, which increases the importance of transliteration problem for the next generation. Each of those transliterated materials becomes a valuable resource for the preservation of the national identity. However, due to the numerous changes made to the scripts, various transliteration versions have been generated and spread on the web. Solution of this problem requires the adoption of a unique transliteration standard for Azerbaijani language, including the adaptation of social and psychological and moral thinking of Azerbaijani people.

Table 1. Modified ISO Romanization table for Azerbaijani language used for the system

As mentioned above, the spread of various transliteration forms leads to confusion. This problem exists almost in all languages. For example, Azerbaijani last name «Quliyev», which has numerous transliterations as Kuliyev, Quliyev, Guliyev, Kuliev, Quliev, and Guliev. The reason for these variations is due to the long-term impact of Russian language of Azerbaijani language. Thus, the letter «Q», equivalence of which in accordance with the phonological features of this letter in Azerbaijani language is «G», was incorrectly transliterated as «K» into Russian language. Since the letter combination «ye» sounds same as the Russian letter «e», it was transliterated into «e», omitting the letter «y». Consequently, the identity of numerous personal names and place names is violated.

At present, transliteration issue is closely related to the multilingual nature of the global web, in which each of existing and endangered languages tries to be properly represented and not to be pressed by others, and to preserve the national identity. Many organizations dealing with the transliteration tables do not own transliteration systems, which can process different scripts. One of the reasons of this fact is insufficient funds, whereas another reason is related to the unwillingness or less concern to meet the multilingual needs of community (Simsova, 1988). Since most of organizations are restricted to the Roman characters and do not ensure transliteration tables for non-Roman scripts, users

cannot find or access the needed resources, materials and documents in their original scripts. Consequently, Romanization becomes urgent being provided with the automated cross-language transliteration system, which comprehensively represent the graphical and phonological features of certain language.

For comprehensive solution of transliteration problems, in this study, Romanization table for Azerbaijani language ISO is modified and provided in the Table 1. The modifications are made by the experts of certain fields of sciences in accordance with the grammatical, graphical and phonological features of Azerbaijani language. Based on this conversion table, an automated transliteration system for Azerbaijani language with other most commonly used languages by the Azerbaijanis, namely Russian, English, French, German and Farsi, is developed. Transliteration system is improved to achieve maximum transliteration accuracy based on feedbacks and proposals submitted to the system by the users.

Simple replacement principle is used for the system transliteration. The advantage of the system is its independence from any handmade corpus as the system is intended not only for the transliteration of proper names and terms but for all words and texts. For perfection of the automated transliteration the feedbacks and errors are regularly controlled, checked and corrected by the experts of the field.

CONCLUSION

The problem of transliteration of Azerbaijani language with other languages, particularly with English becomes topical, taking into account the number of Azerbaijanis living throughout the globe accounting for 30-35 million people (Sela, Avraham, 2002). Given the dominance and the status of international language of English, transliteration from Azerbaijani language into English and vice versa becomes an essential element of their communication with each other via the virtual environment.

This paper explored the growing dominance and status of English as an «international language» and «lingua franca». One of the key elements of communication on the web, that is transliteration, its importance and challenges were highlighted. The requirements for the users to transliterate necessary resources and information written in the modified Latin scripts were explained. Since the web resources include different transliterations of the same words, the factors affecting this situation were identified. This paper focused on and offered some strategies that could contribute to eliminate transliteration problems, preserve national identity and improve information retrieval.

REFERENCES:

1. Crystal, D. (2003). English as a Global Language (Second edition). Cambridge: Cambridge University Press.
2. Firth, A. (1996). The discursive accomplishment of normality. On «lingua franca» English and conversation analysis. Journal of Pragmatics 26: 237–59.
3. Ammon U.,, (1992), Gengo-to Sono Chii. Trans. by Y.Hieda and H.Yamashita Tokyo; Sangensha
4. Sunol, R., & Saturno, P. (2008). Challenge to overcome language barriers in scientific journals: Announcing a new initiative from the ISQua journal [Editorial]. International Journal for Quality in Health Care, 20(1), pp. 1–2.
5. Nida, E. A., & Taber, C. R. (1969). The theory and practice of translation. Leiden: EJ Brill
6. Library and Archives Canada, Ottawa, (2006) https://www.bac-lac.gc.ca/eng/services/cataloguing-metadata/Documents/040006-14-e.pdf
7. Bedii Duru Altuğ, (2014) The 1991 International Contemporary Turkic Alphabets Symposium and its Contributions to the Turkic Alphabet Reform, University of Washington, pp.8-11
8. B. Weinberg, (1974) "Transliteration in Documentation," Journal of Documentation 30: pp. 18-31.
9. Law of the Republic of Azerbaijan on the renewal of the Azerbaijani alphabet with Latin graphic (December, 25, 1991).
10. Sylva Simsova, (1988), "Coping with Foreign and Nonstandard Character in Libraries," in John Eyre, ed., Small Computers in Libraries, London: Meckler, pp.13-15
11. Sela, Avraham (2002). The Continuum Political Encyclopedia of the Middle East. Continuum. p.197. ISBN0-8264-1413-3.