

International Conference on Turkic Languages Processing

(TURKLANG2021)

**Application of models for transliteration of the Azerbaijani
scripts with the scripts of other languages**

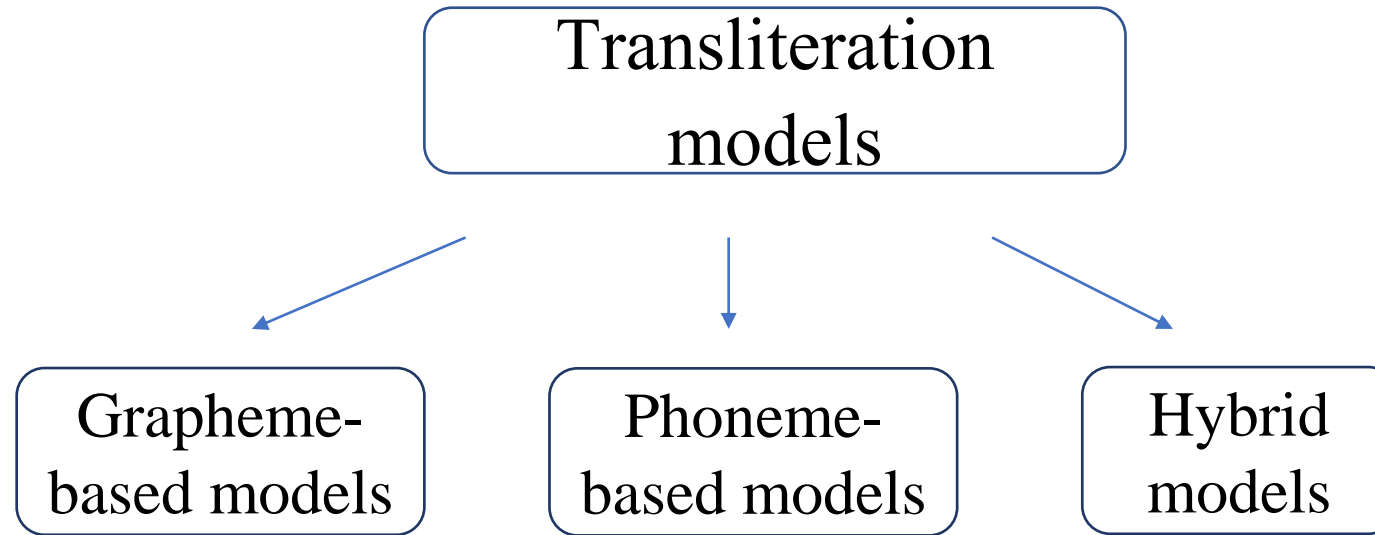
Sabina Mammadzada

Senior researcher

Institute of Information Technology of
Azerbaijan National Academy of Sciences

September 21-23, 2021

1. Types of transliteration models



1.1. Grapheme-based models

A grapheme is a group of characters belonging to the same alphabet.

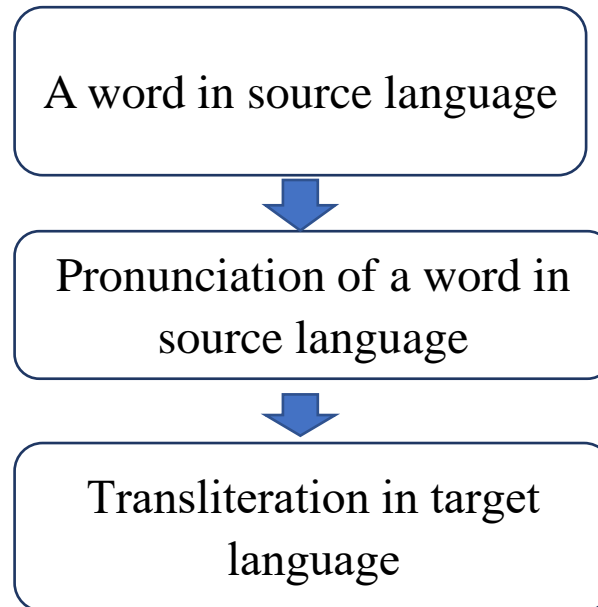
In grapheme-based approaches, transliteration is performed by directly converting the graphemes of the source language into the graphemes of the target language, without taking into account the phonetic features of the language.

- Statistical Machine Transliteration model,
- Rule-based models,
- Hidden Markov Model,
- Finite State Transducer, etc.

1.2. Phoneme-based models

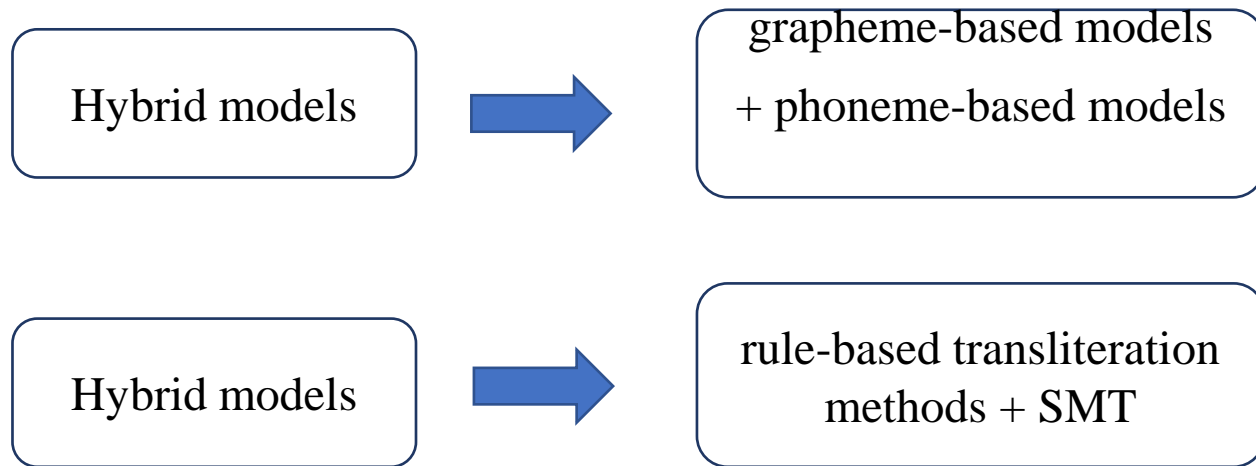
A phoneme is the smallest meaningful unit of sound or the smallest contrasting (distinguishing) unit of language.

In phoneme-based models, phonetic features are taken into account in the transliteration process, rather than the spelling features. Phoneme-based methods are also known as the Pivot method.



1.3. Hybrid models

Hybrid model is a combination of two or more transliteration approaches is required to achieve more accurate transliteration. Hybrid approaches combine grapheme-based transliteration models with phoneme-based transliteration models. The hybrid transliteration approach uses statistical and rule-based transliteration methods



2. State-of-the-art in the field of transliteration of the Azerbaijani language

- The Latin scripts adopted for the Azerbaijani language differ from those adopted for English, French, German and other popular languages, both in terms of sounding and quantity;
- Problematic letters, such as ç, ğ, ı (lower case for letter “I”), İ (upper case for letter “i”), ö, ş, ü) and (ə);
- Some letters adopted for different conversion tables for the Azerbaijani transliteration as ž, j, ě, ï, ^, ù, š, č, ê;
- Some letters do not exist in the Azerbaijani language, as w.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.1. For Azerbaijani-Russian transliteration:

In the Azerbaijani alphabet, depending on the position of the letter ə, it is replaced by the letters *a*, *e*, *я* and *э* according to the following rules:

- at the beginning of the word and after the letters *b*, *s*, *z*, *r*, *n*, *ş*, *q*, *ç*, *c*, *j*, *f*, *x* and *h*, in Russian it is often replaced by the letter *a*. For example, *Nəriman* – *Нариман*;
- at the beginning of the word and after the consonants *k*, *g*, *l*, it is replaced by the letter *я*. For example, *Ləki* – *Ляки*;
- it is replaced by the letter *e* when it comes after the consonants *b*, *v*, *ğ*, *d*, *z*, *p*, *r*, *s* and *t*. For example, *bəy* – *бей*;
- At the beginning of the word, it is often replaced by letters *э*. For example, *ət* – *эм*.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.2. For Azerbaijani-English transliteration:

In the Azerbaijani alphabet, depending on the position of the letter *ə*, it is replaced by the letters *a* and *e* according to the following rules:

- in single-syllable words before the letters *l*, *m*, *n*, *r* and *y*, and in two-syllable words before and after the letters *l*, *m*, *n*, *r*, *y* and *z*, it is replaced by the letter *e*. For example: *Azər* - *Azer*.
- in other cases, it is replaced by the letter *a*. For example, *Gəncə* - *Ganja*.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.3. For Azerbaijani-German transliteration:

- The letter *s* in the Azerbaijani alphabet is replaced by the letter combination *ss* in the first syllable of the word to be pronounced as *s* before the letters *p* and *t*, for example, *stol* - *sstol*.

3.4. For Azerbaijani-French transliteration:

The French equivalent of the letter *ı* in the Azerbaijani alphabet corresponds to the letters *i* and *y* according to the following rules:

- The vowel *ı* before the closed syllable is replaced by the letter *y*. For example, *Laçın* - *Latchyn*.
- In most other cases, it is replaced by the letter *i*. For example, *Xızı* - *Khizi*.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.5. For Azerbaijani-Italian transliteration:

The Italian equivalent of the letter *z* in the Azerbaijani alphabet corresponds to the letters *z* and *s* according to the following rules:

- in some cases, it is replaced by the letter *s* in the middle of the word, between two vowels. For example, *Alazan* - *Alasan*.
- in other cases, it is replaced by the letter *z*. For example, *Zinəddin* - *Zinedine*.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.6. For Azerbaijani-Spanish transiteration:

The Spanish equivalent of the letter *g* in the Azerbaijani alphabet is replaced by the combination of the letters *gh* and *gu* according to the following rules according to the position in the word:

- It is replaced by the letter combination *gu* before the letters *e* and *i*. For example, *Gədəbəy* - *Guedebeý*.
- In other cases, it is replaced by the letter combination *gh*. For example, *gül* - *ghul*.

3. Application of models for transliteration of Azerbaijani scripts with the scripts of other languages

3.7. For Azerbaijani-Farsi transliteration:

The letters *e*, *i*, *o*, *ö*, *u* and *ü* in the Azerbaijani alphabet:

- in Farsi alphabet at the beginning of the word in Farsi alphabet they are replaced by letter combinations *e* - *ای*; *i* - *ی*; *o* - *او*; *ö* - *او*; *u* - *او*; *ü* - *او*.
- in the middle and at the end of the word in the Farsi alphabet they are replaced by letters *e* - *ی*; *i* - *ی*; *o* - *و*; *ö* - *و*; *u* - *و*; *ü* - *و*.

Thank you for your attention!