

## Issues in Udi Orthography

Caucasian languages are renowned for their phonological complexity, including large segmental inventories, long words, morphophonemic alternations, and variation between dialects. The size of the segmental inventories has been particularly problematic, requiring orthographies making extensive use of diacritics or diphthongs. In this paper we examine how the segmental inventories have been represented in various orthographies of the Udi language. In particular, we outline the role of linguistic, sociolinguistic and political factors in a newly proposed orthography currently being used in language development activities.

The Udi language is a member of the Lezgi family of North Caucasian languages. Traditionally the Udi people lived in the villages of Oğuz (formerly Vartaşen) and Nic in the Qəbələ District of north-central Azerbaijan. Most speakers of the Vartaşen dialect now live in Georgia. Our research is focused on the Nic variety of Udi.

The phonological system of Udi has 15 vowel phonemes and 35 to 38 consonant phonemes depending on the analysis. The Cyrillic orthography used in early Udi primers was based on the technical orthography used by Gukasjan (1974). It contained 15 graphemes for vowels, and 37 graphemes for consonants.

Although the Cyrillic orthography was used in the schools, and a fair amount of literature was published using it, few people ever became proficient in using it. We present a number of factors that undoubtedly contributed to this. First, of the 52 total graphemes, 24 were digraphs and 2 were trigraphs. This made long words even longer, and therefore presented problems for developing word attack skills. For example, /ædʒyɣhaq'sun/ is written as <аьджубьгьгьақьсун>. Second, most of the digraphs and trigraphs were built using the hard sign <ъ>, the soft sign <ь>, or the 'stick' <I>. The similarity between the two modifiers <ъ> and <ь> adds to the difficulty of phoneme identification. Furthermore, these modifying elements were not used consistently. Finally, the system did not make use of graphemes from the Azerbaijani orthography for phonemes that do not in Russian. For example, instead of using the Azerbaijani <ƒ> for /ɣ/, the Udi system used <гъ>. This meant the system did not use all the transferability that was available.

The adoption of the Latin alphabet by Azerbaijani presented an opportunity to change the orthography. In the mid-1990s, a Latin-based orthography was developed for a new set of primers. This new system used the Azerbaijani graphemes for phonemes that do not occur in Russian, increasing transferability. In spite of this, the new system did not result in greater acceptance of Udi literacy. As we demonstrate, a number of the factors that contributed to this were the same as those already discussed in relation to the Cyrillic orthography. For the many phonemes that occur in neither Russian nor Azerbaijani, the new system still relied on digraphs. In addition, the system was a mixture of Cyrillic and Latin, using basic shapes like <ц> and the modifiers <ъ> and <ь>. This was disliked both by readers fluent in Russian and those fluent in Azerbaijani. Finally, infelicitous choices were made for some phonemes, resulting in orthographic forms like <бьбьбьбь> for /b b / 'bridge'.

In light of the problems with previous orthographic systems, we decided to develop a new system, building on previous systems, but avoiding the problems of the past. We felt the following points

were important to consider in this attempt:

1. Minimize the number of phonemes represented;
2. Use only Latin graphemes;
3. Whenever possible, base the orthography on Azerbaijani;
4. Use diacritics instead of digraphs where possible;
5. Use diacritics consistently.

The system we have proposed meets all but point 1. Thus, transfer from Azerbaijani is maximized, words are shorter, and there is less confusion between graphemes.

The first point was explicitly mentioned by a group of Udi speakers involved in language development who felt that one major reason for the difficulties they and others had with the previous orthographies was due to the large number of sounds represented. With so many graphemes, they were finding it difficult to spell consistently. We hoped to eliminate as many as ten graphemes on the basis of previous work that indicated that a number of sounds were phonologically conditioned or carried low functional load. When we shared our recommendations with those involved in development, however, they decided that all but three of the graphemes were needed, and that the basic problem was one of training, not one of too many graphemes.

The system outlined in this paper is now being used in Nic, apparently with positive results. The next step will be to develop primers so the new system can be taught in the schools. In addition, if this system continues to prove itself, it may point the way forward in developing Latin-based orthographies for other Caucasian languages spoken primarily in Azerbaijan.

#### References

1. Gukasjan, Voroshil. 1974. *Udinsko-azerbajdzhansko-russkij slovar'*. Baku: Elm.