Lexicographical Problems in Written Modern Uyghur

By WOLFGANG-E. SCHARLIPP, Copenhagen

Introduction

Modern Uyghur or New Uyghur or Eastern Turki as it was called some time ago and as GUNNAR JARRING – the most eminent scholar on this language – preferred to call it, is spoken and written by a number of people that can only be estimated. The round number of 8 mio comes probably near to reality. Over 90% of them live in the Western-most province of China, i.e. Xinjiang (Sinkiang), or as we prefer to say: Eastern Turkistan, i.e. Doğu Türkistan.

Eastern Turkistan is inhabited by a number of other nationalities, which partly have a minority status and of which most speak another Turkic language. During this century the demographic situation has changed dramatically in that country, due to planned sinisization, as we also know it from Tibet, where it has resulted in almost the complete destruction of the indigenous culture.

The effect that the demographic development had was not that strong on the Uyghurs, although a look at statistics has something alarming for the future.

According to the official censuses in the year 1940 there were 2,941,000 Uyghurs and 294,000 Han-Chinese living in Sinkiang. In 1982 the numbers were 5,949,664 Uyghurs, and 5,286,533 Chinese. While the number of the Uyghurs, and by the way also of the Kirgiz and Özbek has doubled, that of the Kazaks even tripled, Han-Chinese have become 18 times more within these 40 years. Thus it does not take wonder when many places in Sinkiang make rather a Chinese impression instead of a Turkic one, except that labels and posters are usually written in Chinese and in Uyghur.

One of the highest percentage of Chinese inhabitants of course is to be found in the capital Ürümqi with 75,62%, but even in smaller places the Chinese are now in the majority, so for example in Komul/Hami, where they make 68,09%.
The Modern Uyghur Written Language

The aforementioned numbers suggest a constant influence of the Chinese language upon Modern Uyghur. In this context we shall come back to Chinese loanwords a bit later. It should be mentioned that there is also an influence of spoken Uyghur on spoken Chinese in Eastern Turkestan, of course to a lesser degree, as the percentage of bilingual Chinese speakers is much smaller than that of bilingual Uyghurs. Anyway Uyghur loanwords in the Chinese language of Eastern Turkestan are restricted to goods or traditions typical for Uyghurs, like for example nan for "bread".

Despite Chinese influence Uyghur is a language with a rich literary development during this century, comprising all kinds of literature, belle-lettre and science. On the next pages we shall cast a short glance at the development of the orthography of Modern Uyghur, as some of the problems for lexicography arise from the orthographical system while others have their source in their semantic complexity.

Writing systems for Modern Uyghur

As it is widely known, most of the Central Asian Turkic languages, also Uyghur has been written in different alphabets. During the first two decades the written literature that could be found among Uyghurs was written in the Arabic alphabet using in principle Chagatay orthography. A short look at the main characteristics will suffice for our purpose.

Concerning phonetics

Sounds of the Turkic language that did not have an Arabic letter to represent them were – like in Ottoman – often written by a sign that gave nothing but a clue to the Turkish linguistic value, like f. ex. the letter вав standing for five sounds with phoneme-value.

Concerning consonants, the guttural nasal /ng/ was written with a digraph n + k according to the tradition of Classical pre-Islamic Uyghur. I assume that these facts are known to the reader, thus we don't have to go into detail in this respect.

The graphical representation of consonantal assimilation is highly neglected. In all we can state that the ruling principle of organizing linguistic entities i.e. orthography is based on morphological principles, in the case of Arabic and Persian loanwords additionally on etymological principles, as they are written like in the respective languages and not according to Uyghur pronunciation.
Some examples

- Vowels were in general written with matres lectionis as in okumak "to read".
- Case suffixes were written separate from the stem of the word as for example in kitab-ni "the book" (Akk.).
- Disregarding velar or palatal quality of the syllable, the same consonant letters in anlaut position are used where in Ottoman different letters are taken, i.e. for example the emphatic variant of the Arabic sibilant in velar surrounding: söz "the word"; salmak "throw, let go". Ottoman has instead صلما. But as in Ottoman in the dental anlaut of the suffix there is no synharmony i.e. the consonantal assimilation is not reflected in orthography: خزم "in writing etc." There are exceptions - even as a rule, which will probably be of graphotactical origin, as for example in قيلغاندانا kilğanda, where the locative suffix is connected with the base.

The modern writing systems

As the chronological development of writing Uyghur is not the topic of this article, we will just mention in short the most important facts. In the year 1923 a scientific commission founded by the "Yettisu Uygur märkäzi Komititi" of the communist party had decided to use the Arabic alphabet for writing Uyghur, according to the model given by the Tatar reformers (see Baldaufl p. 192f.).

Between 1928 and 1938 an alphabet based on the latin alphabet was introduced for the Turkic languages of the Soviet part of Central Asia, also of the Uyghurs in Kazakhstan, while the Uyghurs in Sinkiang continued using the Arabic alphabet. The introduction of alphabets based on Cyrillic during the years after 1938 had as one of its goals the easier influence of Russian on the Turkic population. A similar approach was done by the Chinese in 1959, when a latin-based alphabet was introduced for Uyghur and Kazakh in Sinkiang, including the pinyin-letters, which facilitated the introduction of Chinese words in written Uyghur. In the following I want to concentrate on some of the principles of orthography, mainly in the Arabic alphabet, but in some cases also to the cyrillic one, still being used for Uyghur in Kazakhstan, and the Latin based one, still being in use by some people, who had their education during the period, in which this alphabet was introduced.
Writing stems

In Uyghur the phonetic law of regressive assimilation reduces a broader \( a \) under the influence of either a following closed vowel in the stem or suffixation by a suffix including a vowel. So we find for example the Turkish word \( agiz \) “mouth” pronounced \( æ̈giz/ɛ̈giz \) or in some similar forms. In modern Written Uyghur standard orthography the anlaut vowel, being a closed \( e \) is written with the Arabic letter \( ya \) and two vertical dots under it, a letter created for this purpose. Also loanwords are treated according to this law, for example when they are suffixed with a suffix beginning with a vowel, as in the case of \( mal \) (“property”) > \( meli \) (“his, her property”).

Another phonetic phenomenon is called vowel reduction. It refers to the fact that \( a/å \) being the last sound (auslaut) of a root changes to \( i \) if a suffix is added containing a vowel. In both cases the change of the stem vowel is represented in writing, which leads in the case of \( barmak \) “to go” and \( bârmâk \) “to give” to homographs, according to the law of regressive assimilation, when the deverbal noun suffix -\( i \)\( š \) is added: \( bar + i \)\( š \) > \( beri\( ş \); \( bâr + i \)\( š \) > \( beri\)\( ş \). Also in the case of vowel reduction this is realized in writing: \( bâtilâr \).

In contrast the new orthography in all three alphabets is now based on phonetic-phonemic principles (not on phonological principles in the sense that only sounds with phoneme value are represented by graphic signs). This can also create minor or major problems for lexicography.

Loanwords

Before dealing with lexicographical problems created by Uyghur words, some short remarks will be made to the problems created by loanwords, as these brought problems with them due to their pronunciation.

While the writing of Russian words in Uyghur created no problems in the cyrillic script, writing Chinese words in Arabic was problematic, as long as there was no normalized way of writing them. The introduction of the pinyin-letters made it possible to write in Uyghur – as well as the other languages of China – Chinese words according to Chinese phonetics. Thus Sinkiang was then written Xinjiang, which was not the way it was pronounced by Uyghurs. By re-introducing the Arabic alphabet, the Uyghur-Turkish pronunciation was given again.

Loss of sounds

One of the problems regarding Uyghur words, is to decide which form of a word is to be included in the dictionary, if several variants of the pronuncia-
tion can be found. One such case is the loss of /r/, a sound which is disappearing in many languages in certain positions of the word, especially in the auslaut-position of a syllable.

Thus for example the /r/ in the end of the stem of the verb bärmäk is mostly not pronounced. In the Uyghur dialect dictionary of G. Jarring, the importance of which for Uyghur studies cannot be underlined enough, we find seven different phonetic entries for this verb, only two of them having the /r/ pronounced, while the stems without /r/ have their own entry with a reference to the entry bärmäk. This is of course the ideal approach in a dialect dictionary, but not advisable in a dictionary of a normed written vernacular.

The authors of the *Uygursko-Russkii slovar’* from 1939 recognized the problem, but their solution goes rather into the direction of a dialect dictionary, as they put the /r/ in those positions where it is more often not spoken, into brackets: actur/rjmaq “to open”. In the later dictionaries, for example Nadžip and St. John, which we shall refer to in the following, only the forms with /r/ are given. This is in accordance to the decision of language planning, which regarded in such cases obviously the etymological way of spelling a word for the better i.e. more exact language.

This etymological principle was not always used. Not in those cases where the loss of /r/ had taken place far back in the past and the resulting pronunciation had become the commonly accepted one. Thus we find in all three quoted dictionaries the word nāgā “where to?”, spelled nāgā in the Arabic, the cyrillic (Nadžip) and the Latin alphabet.

Still in 1934 Denison Ross writes it in Arabic script سمركة, but giving already the phonetic transcription nā-y(ə)r-gā, resulting in the pronunciation nāygā, which is one of six different pronunciations that Jarring gives.

**Haplology**

Another problem belonging into this category is haplology, a phenomenon occurring very frequently in verbal compounds, which are treated from another point of view further down. So for example the converb construction of the verbs almak “to take” etc. in the first and kālmāk “to come” in the second position create the new lexeme “to bring, fetch”. Such common and frequently used words as this one show – as can be expected – a high degree of forms caused by haplology.

G. Jarring gives as many as nine different phonetic forms of this compound, among then alip kel-, apkel-, ekel-, etc. The full form alip kel- has to be regarded as a word of an older, maybe classical and literary language, while the others are words of sociolect and dialect, the most common of the
spoken language being *akel-* Dictionaries handle this problem in different ways. For example NADŽIP includes both forms, saying in the case of *akel-* that it is a contracted form of *alsp kel-* St. John does not have any compound entries under *almak*, but gives the form *ekkelmek* without any reference to its etymological origin. The reason for this is that this form is the one that has become the form generally accepted also for the modern literary language.

Problems of arranging in order

As had already been mentioned, the phonetic principle of spelling causes a change of stem vowels if certain suffixes are added. This leads in the case of arranging the entries of a dictionary in a strict alphabetical order to the problem that different derivations of the same stem have to be put into different places of the dictionary.

Thus the deverbal nouns in -iş of both *bärmağ* and *barmak*, producing homophones and consequently homographs, as we had seen above, are now entries in the same place, but both in a place after the entries of their respective verbal stems, as the closed /e/ has a later place in the new alphabet.

This refers to all similar cases, as for example *almak* “to take, buy” *eliş* “the act of taking, buying” etc. Again dictionaries treat this phenomenon in different ways.

NADŽIP gives under the entry *satmak* “to sell” those verbal compounds which create a new semantic entity: *setip almak* “to buy”, *setip eliş* the verbal noun of this, *setip bärmağ* “to sell”, but gives simple derivation of *satmak* according to the change of the vowel in another place of the dictionary, i.e. *setilmak* “to be sold” and *setiş* (verbal noun), saying explicitly that these are derived forms of *satmak*. St. John has no other entry under *satmak* (p. 223) but gives the passive *setilmak* and the noun *setikçi* “businessman” on p. 245. In the same way the *Uygursko-Russkiy slovar* proceeds, also without referring to the stem having its entry in another place.

According to a strictly alphabetical order in lexicalization, stems and their respective derivations for example like: *yanmak* “to return”, *yeniş* “the return”, *yandurmak* “to recall” have different places in the dictionary. If such is the case, at least there should be a reference to the original stem, in order to make lexical connections clear.

Verbal compounds

Normal verbs functioning as a modifier of a preceding verb ending in a con-verb form are a very common feature in all Central Asian Turkic languages.
In the Southwestern Turkish group these verbs are often called *tasvir fiili* i.e. descriptive verb, which is not a satisfying term for a group of verbs with very different functions.

The function of verbs in immediate post-converb position has been studied by several scholars regarding different Turkic languages. An important question to answer in most of these studies was, into which linguistic category these full verbs in auxiliary function would fit, whether they should be regarded as expressing aspect or aktionsart. The most recent and thorough discussion and summary of this problem is found in C. Schönic’s study of these verbs in the Tatar language, to a certain degree also in Scharlipp’s study of the same topic in Modern Uyghur (1984).

In this place we are concerned with the semantics of these verbs only in so far as their lexical status is concerned.

**Verbs between lexemes and morphemes**

Two verbs of duration character have become part of certain tense paradigms: *turmak* “to stand upright, come to a halt etc.” and *yatmak* “to lie down, to be lying”, while also having the status of a full verb.

There is no general agreement about the participation of *turmak* in the various tense paradigms, as in some of them there are only traces of *turmak* in the forms of the suffixes -ti, -tu and -du, and usually only in one or two persons of the same paradigm. Thus for example in the so called past narrative tense we find the first person singular *yeziptiman* probably a contracted form of *yezip turur man*, no rest of *tur* in the forms of the second person *yezipsan* and *yezipsz*, where *tur* has obviously been completely dropped due to conflicts in Uyghur pronunciation plus frequency of the use. The form of the third person offers the clearest evidence for *turmak* as a part of the paradigm, as it is *yeziptu*.

There is a similar situation in the continuous past tense 1. *yeziwatiptiman*, 2. *yeziwatipsan*, *yeziwatipsiz*, 3. *yeziwatiptu*. In the paradigm of the normal present tense the situation is even unclearer, as a rest of *turmak* appears in the third person only: *yazidu* “he, she, it writes”. A study in the history of these forms and the probability of *turmak* being involved in those forms where no trace of it is left was published by L. Johanson 1976.

As *turmak* is also used as a main verb, it has to be regarded as a lexeme, giving all its meanings it can have as such a verb. Its function in the above mentioned tense forms can only be reconstructed etymologically and thus have their place in the grammar, where they have to be treated as suffixes, but if possible with a commentary as to their history.
Main verbs in auxiliary function

The situation is quite different, when using *turmak* after a converb, which has kept its recognizable form. It can firstly express the continuity of the action expressed by the converb preceeding it. In this function it can be freely used with any verb, which will by its nature mostly be a verb which expresses anyway an act continuous by nature. In this case it produces an "aktionsart" and thus does not change the semantics of the preceding verb.

In its second function it does exactly do this, it modifies the preceding verb according to the semantics of that verb. So for example after the converb of *çikmak* "to leave, get up etc.", it does not have the function of creating a durative, but it adds to the ingressive or inchoative action of *çik* - its own meaning of "standing upright" thus producing the compound *çikip turmak*, meaning "to stand up, get up".

A case which is especially misleading, looking at it from the Turkish perspective, is the verb *bärmäk* which of course corresponds to Turkish *vermek*, but which gives after the converb in -p the information that the action expressed by the preceding verb happened or was carried out in favour of someone else and not for one self. For this reason the term "parasmaipada" should be considered adequate, as this is the term for the same function of a verbal category in Sanskrit.

Thus the compound *okup bärmäk* does not correspond to Turkish *oku-yuvermek* "to read quietly" or the like, as we find it translated in the Turkish edition of NADŽIP's dictionary (see under NECIP in the list of literature), where many of such misunderstandings concerning Uyghur verbal compounds can be found. It would have to be translated into Turkish *birine bir sey okumak*. The Uyghur verb *bärmäk* has in this case the function of the German prefix vor- in the verb vorlesen, reading something to someone. These false translations into Turkish are the more surprising as NADŽIP himself renders the correct Russian equivalents.

The question arising from this situation is, which of the functions should be included in a dictionary and which ones belong into a grammar?

A. MArtINET’s pessimistic opinion (of the year 1970) about the possibility of establishing criteria for the definition of a “word” is certainly still valid:

"Il serait vain de chercher à définir plus précisément cette notion de mot en linguistique générale. On peut tenter de le faire dans le cadre d’une langue donnée. Mais, même dans ce cas, l’application de critères rigoureux aboutit souvent à des analyses qui ne s’accordent guère avec l’emploi courant du terme.” (p. 115)

Nevertheless, according to our opinion, the lexicographer, being confronted with a clear semantic entity designated by a certain phonetic shape, has to
decide according to the criteria of his method. In my opinion NADŽIP had already shown the right way in lexicography, by including as lexemes all those compounds which produce a verbform in which the semantic compounds of both verbs have merged to a new semantic entity, i.e. an objective change in the quality of the action.

On the other hand those verbs, which produce in their post convert verb position systematically such a change should also have their place in the grammar, as their grammatical function cannot be seen different from that of a suffix.

There are two possibilities to handle the placement of semantically identical or coherent words in a dictionary. The one is, making rich use of reference to the stem, which is in another place. This is done to varying degrees in the existing dictionaries. The second solution would be the more difficult one, i.e. a change in the orthography, by introducing an etymological principle, as the problems we face, are caused by an exaggerated phonetical system, which was introduced during the first two decades, of the last century when this idea of “writing as close to the spoken language as possible”, was regarded as being the ideal principle of orthography. A lot of other ideals have been discussed since then.

References


—: Studien zu einer osttürkischen Lautlehre. Lund 1933.


---: Türkische Sprache – Arabische Schrift. Budapest 1995 (BO. XLIV.)
Sima Qians Huo-Zweifel in Kapitel 61 des Shiji

Von DOROTHEE SCHAAH-HANKE, Hamburg


Bei der Suche nach Wörtern für „Zweifel“, die Sima Qian explizit mit Bezug auf sich selbst in seinen persönlichen Bemerkungen verwendet, stößt man neben dem Wort yi 疑 an einer einzigen Stelle des Shiji auch auf das Wort huo 漢. Ulrich Unger übersetzt dieses spezielle Wort zur Bezeichnung von Zweifel mit „Verwirrung“, „Verirrung“ und fügt erläuternd hinzu:

Orientierungslosigkeit, Unfähigkeit, logisch zu denken oder sachgerechte Entscheidungen zu treffen, manchmal auch einfach Zweifel, Unsicherheit. Hier und da steht huo für „Irrtum“, also für das Verkennen der Wahrheit.  


---

1 Für zahlreiche Anregungen und Korrekturen zu früheren Versionen dieses Beitrags bedanke ich mich sehr herzlich bei Prof. Dr. Hans van Ess, Prof. Dr. Stumpfeldt und Prof. Dr. Michael Friedrich.