

# THE TURKIC LANGUAGES IN A NUTSHELL

## The Internal Classification & Migration of Turkic languages

Version 8.1

**v.1** (04/2009) (first online, phonological studies) > **v.4.3** (12/2009) (major update, lexicostatistics added) >

**v.5.0** (11/2010) (major changes, the discussion of grammar added) > **v.6.0** (11-12/2011) (major corrections to the text; maps, illustrations, references added) > **v.7.0** (02-04/2012) (corrections to Yakutic, Kimak, the lexicostatistical part; the chapter on Turkic Urheimat was transferred into a separate article; grammatical and logical corrections) > **v.8** (01/2013) (grammatical corrections to increase logical consistency and readability, additions to the chapter on Uzbek-Uyghur, Yugur)

### Abstract

The internal classification of the Turkic languages has been rebuilt from scratch based upon the phonological, grammatical, lexical, geographical and historical evidence. The resulting linguistic phylogeny is largely consistent with the most prevalent taxonomic systems but contains many novel points.

### Contents

#### 1. Introduction

#### 1.1 Preliminary notes on the reconstruction of Proto-Turkic

#### 2. Collecting factual material

#### 2.1 An overview of the lexicostatistical research in Turkic languages

#### 2.2 Dissimilar basic lexemes in the Turkic languages

#### 2.3 The comparison of phonological and grammatical features

### **3. Making Taxonomic Conclusions**

**Bulgaric**

**Some of the exclusive Bulgaric features**

**Yakutic**

**Where does Sakha actually belong?**

**How did Sakha actually get there?**

**On the origins of Turkic ethnonymy**

**Altay-Sayan**

**Tofa and Soyot closely related to Tuva**

**The Khakas languages**

**Khakas and Tuvan share no exclusive innovations**

**Altay, Khakas and Tuvan form the Altay-Sayan subgroup**

**Great-Steppe**

**Kimak-Kypchak-Tatar, Kyrgyz-Kazakh, and Chagatai-Uzbek-Uyghur seem to form a genetic unity**

**Great-Steppe and Altay-Sayan seem to be closer to each other than to Oghuz-Seljuk**

**Kyrgyz-Chagatai**

**Kazakh is closely related to Kyrgyz**

**Altay-Kyrgyz isolexemes**

**Chagatai looks like Karakhanid affected by Kyrgyz**

**Kimak-Kypchak-Tatar**

**The Kimak subtaxon**

**The relationship between Oghuz and Kimak**

**On the origins of the ethnonym Tatar**

**Bashkir is closely related to Kazan Tatar**

**On the origins of Nogai**

**Karachay-Balkar, an atypical Kimak language**

**Oghuz-Seljuk**

**Oghuz is still a valid subtaxon**

**Seljuk as a subtaxon of Oghuz**

**Oghuz-Seljuk is indirectly related to Orkhon-Karakhanid**

**Notes on the confusion about y-/j- in Oghuz and Kimak**

**Orkhon-Karakhanid**

**Orkhon-Karakhanid as a valid subtaxon**

**Khalaj is probably an offshoot of South Karakhanid**

**Yugur-Salar**

**Yugur seems to be ancient**

**Salar has little to do with Oghuz, but quite a lot with Yugur and Uyghur**

**4.The Resulting Internal Classification of Bulgaro-Turkic languages**

**4.1 The Genealogical Classification of Bulgaro-Turkic languages**

**4.2 The taxonomic Classification of Bulgaro-Turkic languages**

**4.3 The Geographical Tree of Bulgaro-Turkic languages**

## 5. References and sources

### 1. Introduction

The present study of the Turkic languages (2009-2012) was started as brief online notes that gradually grew into a series of online publications. The study is mostly an original research with relatively few references to previous theories. Most analysis was based upon *factual evidence* collected from dictionaries, grammars, language textbooks, native speakers on the web, sound and video fragments, books and articles containing detailed descriptions of specific languages. The resulting conclusions rarely draw from historically accepted opinions or assumptions produced by other researchers, rather attempting to build a logically consistent view of the spread of Turkic languages and their internal classification grounded in the nearly independent and relatively comprehensive step-by-step analysis.

Nevertheless, the author deeply appreciates the extensive input from people who worked on the vast amount of Turkological literature dedicated to the numerous Turkic languages, as well as those who helped directly or indirectly by providing corrections and valuable notes by email or through web forums, without whose interest and collaboration this work would never have come to life.

The present article provides all the linguistic argumentation concerning the *internal* classification of Bulgaro-Turkic languages. Furthermore, there are three other separate articles which can be regarded as part of the same work.

*The Lexicostatistics and Glottochronology of the Turkic languages (2009-2012)* is a detailed research of Swadesh-210 wordlists, which dates the Turkic Proper split to about 300-400 BC, and the Bulgaro-Turkic split to about 1000 BC.

*The Proto-Turkic Urheimat & The Early Migrations of the Turkic Peoples (2012-13)* is a detailed analysis of the early Bulgaro-Turkic migrations largely based upon the results obtained in the glottochronological analysis above and the present classification. The Proto-Turkic Proper Urheimat area was positioned northwest of the Altai Mountains, and the earlier Proto-Bulgaro-Turkic Urheimat in northern Kazakhstan. The work explores the associations with the major archaeological cultures of the Bronze and Iron Age period in West Siberia.

*The Turkic languages in a Nutshell (2009-2012)* embraces the final classification, trying to focus on the most well-established conclusions from various works including the present investigation. It also contains multiple illustrations, notes on history, ethnography, geography and the most typical linguistic features, which essentially makes it a basic introduction into Turkology for beginners.

### 1.1 Preliminary notes on the reconstruction of Proto-Turkic

Before we proceed with the main analysis, let us consider the reconstruction of the Proto-Bulgaro-Turkic word-initial *\*j/\*y*, which has become a long-standing issue in Turkological studies, and which may affect certain conclusions in the main part of this publication.

Many proto-language reconstructions in various branches of historical linguistics are often based entirely on the supposed readings of the ancient texts from the oldest family representatives. For instances, in the Indo-European studies we can avail ourselves of the wonderful attestations of Ancient Greek, Latin and Avestan. However, when the oldest representatives are poorly read and interpreted, such an approach can result in errors.

Generally speaking, an ancient extinct language can only be seen suitable for reconstruction purposes, only if it meets several conditions, namely: (1) it is a uniquely preserved language *closely related to a proto-state* without the existence of any alternative sibling branches; (2) it is so *well-attested* that its data are completely reliable and no significant misinterpretations can occur from occasional mistakes in ancient writing, reading (e.g., from abraded petroglyphs), copying of the material, translation, interpretation, etc; (3) the script *closely and adequately reflects the original pronunciation* and we know full well how to correctly reconstruct that pronunciation from that script; (4) the linguistic material should be *dialectally uniform*, in other word it should constitute just one language, not a mixture of various dialects or languages gathered by numerous contributors during generally unknown periods or from unknown areas [which is referred herein as *the Sanskrit dictionary syndrome*].

Obviously, the situation in Turkology does not meet these criteria. Orkhon Old Turkic, the oldest Turkic language attested in the inscriptions from Mongolia, fails to meet the first point (see details below), it barely gets in with the second one, and raises many objections with the third one. In other words, Orkhon Old Turkic may just be *insufficiently old* or much too geographically off-centered to be considered close enough to the proto-state. Moreover, there may be just *not enough correctly interpreted material* for the solid attestation and interpretation of ancient phonology. Orkhon Old Turkic is not as well reconstructed as, say, Latin and Greek in the Indo-European studies, so many readings are quite ambiguous. And finally, it often gets mixed in literature with Old Karakhanid, Old Uyghur and generally unknown Old Yenisei Kyrgyz dialects (given that not all of the Old Turkic inscription were made in Mongolia). Therefore one should not confuse the methodological basis established for the Indo-European reconstruction with the methods convenient for other language branches, such as Turkic. *An old language is not always just good enough.*

As a result, the reconstruction of Proto-Turkic should be conducted by means of a completely different approach, namely using materials from the *well-attested modern representatives* of Turkic languages. *In that case, we should build a reconstruction using a lineal formula with separately determined lineal coefficients representing contributions for each particular language branch.* This method is drastically different from the old-fashioned old-language-for-all model. As an example, when reconstructing Bulgaro-Turkic, we could roughly assign about 50% to Chuvash and about 50% to Proto-Turkic Proper, and then more or less equally divide the second half among the most archaic representatives from the main

branches, e.g. (1) Proto-Sakha, (2) Proto-Altay-Sayan + Proto-Great-Steppe, and (3) Proto-Oghuz-Orkhon-Karakhanid , hence each one of the main Turkic branches would receive only about  $50\% / 3 = 17\%$  (see the classification dendrogram at the end of this article).

This example has been provided as a first-approximation approach to address the potential Old-Turkic-centric attitude, which supposedly claims that "nothing that's not in Old Turkic could exist in Proto-Turkic" or that "Old Turkic is an ancient language, therefore it is more suitable for historical reconstruction". By contrast, the current revised method requires that Göktürk Old Turkic be considered as just one of several early Turkic branches, and it is hardly any more important for reconstruction purposes than about 17% or less.

However, the figures for the lineal coefficients depend on the genealogical topology of the most basic shoots in the internal classification dendrogram. Therefore, using Turkic languages as an example, we come to a general conclusion that *a consistent internal tree-like language group classification must be built before proceeding with the reconstruction of a proto-language*. In other words, an internal classification should be constructed prior to further linguistic or geomigrational analysis.

### **An example from the Revised Model: the reconstruction of the Proto-Bulgaro-Turkic \*S-**

The above reasoning can be exemplified by the following reconstruction of **the Proto-Bulgaro-Turkic \*S-** (the S-symbol should be seen herein as just an arbitrary way to designate the \*y-/ \*j-phoneme as in Turkic *yer / jer* "place, earth", *yol /jol* "way", etc ). A very common error resulting from the Turkish-for-all or Karakhanid-for-all model is the conclusion that the words with the y- were pronounced exactly the same way in Proto-Bulgaro-Turkic. This idea is very common even among Turkologists outside Turkey, and seems to go as far back as the Mahmud al-Kashgari's classical *Compendium of the Turkic languages* (1073).

**Note:** Before proceeding with the further argumentation, we should confine ourselves only to the material *internal* to the Turkic languages, the Altaic and Nostratic languages being a completely separate issue that cannot be regarded herein at any length. This method can generally be called as *an internally-based reconstruction* vs. *full* reconstruction.

**Note:** We try to consistently use the Anglophone-based transcription throughout all the articles as opposed to the German-based transcription that goes back to the 19th century's tradition, therefore /y-/ denotes a semivowel as in "year" and /j-/ or /J-/ an affricate as in "Jack". To avoid occasional confusion, the capital denotation /J-/ has been used in some places for additional emphasis. The digraph /zh/ or monograph /ž/ are approximately similar to the voiced sibilant in French "je" or English "pleasure", "treasure". The use of complex UTF signs was avoided for reasons of readability and technical compatibility. For further details on transcription see *The Turkic languages in a Nutshell*.

The following table summarizes the pronunciation of the Turkic \*S- in the most important branches:

The Reconstruction of the Proto-Bulgaro-Turkic *S		
Subgroup	Phoneme	Remarks
<b>Bulgaric</b>		
Dunai-Bulgar, Kuban-Bulgar	<b>d'; zh-/ch-;</b> <b>j'-/ sh'-</b>	The Dunai-Bulgar texts were written in Cyrillic, though their originals had possibly been written in Greek.  The Bulgaric words in Hungarian are written with the digraph , which should be read as /J-/ (as in Italian that provided basis for the orthography) (see Rona-Tash, and A. Dybo). Some of the Hungarian words have the initial <i>sh-</i> , such as <i>shel</i> ( <i>shelet</i> ) "wind" (cf. Chuvash <i>s'il</i> ). Also, cf. the borrowing <i>zhenchugê</i> "pearls" into Old Russian (attested in 1161) and <i>gyongy</i> into Hungarian.
Chuvash	<b>s'-</b>	palatalized, soft
<b>Turkic Proper</b>		
Yakut, Dolgan	<b>s-</b> , <b>s- &gt; h-</b>	Aspirated between vowels, hence /h/ in Dolgan due to the Evenk substratum.
Tuvan, Tofa	<b>ch'-</b>	slightly palatalized
Khakas, Shor, Chulym	<b>ch'-, n'-</b>	slightly palatalized;

		sometimes an irregular /n-/ before /-i, -i/
Kumandy (North Altai)	<b>ch'-, n'-</b>	as in Khakas
Standard South Altai	<b>d'-/ j-</b>	a palatalized soft /d'/ in writing, though pronounced much like English /j-/, maybe just shorter and with more palatalization.
Karakalpak, Kazakh, Kyrgyz	<b>zh- &lt; j-</b> (west to east); <b>j-</b> (Kyrgyz)	<p>An English-type /j-/ affricate in the eastern dialect of Kazakh probably due to the contact with the Altai-type /d'-/, but a /zh-/ sibilant in the western dialects apparently due to a contact with y-type languages.</p> <p>Although at least one speaker suggested that /j-/ (the voiced /ch-/) was in fact original even in central Kazakhstan, whereas /zh-/ developed in the course of the 20th cent. due to a Russified spelling and pronunciation. That can be true in some cases due to mass bilingualism in Kazakhstan.</p> <p>Similarly, this suggestion is partly corroborated in Melioransky's textbook of Kazakh (1894), who wrote that this sound would be similar in pronunciation to the Russian /dzh/ with "a weak beginning", whereas "the pre-sound ("d") entirely disappears in the western part of the steppe". Consequently, */j-/ rather than /y-/ is reconstructed for the early Kazakh.</p> <p>Also, note /J-/ but /-VzhV-/ between the vowels;</p> <p>An English-type /J/ in Kyrgyz</p>
Kazan Tatar and most other Kimak-Kypchak	<b>j'-</b> before <b>-e, -i</b> <b>y-</b> before <b>-a, -o, -u</b>	Many Kimak-Kypchak languages may have been influenced by the written Kazan Tatar standard in the course of the 20th century, whereas speakers often report a /j-/-type affricate in their native dialects.



		<p>E.g., a speaker of Kazan Tatar insists that his dialect (South Eastern Tatarstan) has a soft /j-/and /y-/ in an allophonic distribution.</p> <p>Al-Kashgari (1072) reports /j-/ for Kypchak.</p>
Ural Tatar	<b>j-</b>	<p>The Ural Tatar is a poorly researched dialect located in the Urals, presumably a result of the Kazan Tatars immigration from the 15th-16th to the 19th centuries and thus retaining the early characteristics of Kazan Tatar.</p>
North Crimean Tatar	<b>j-</b> , sometimes <b>y-</b>	<p>Mostly, always /j-/ in the northern (steppe) dialect, though /y-/ in numbers and a few other common words (such as <i>yaxshi</i>), probably due to borrowings at marketplaces.</p> <p>Moreover, a /j-/ is reported in Yevpatorian Crimean Tatar.</p>
Karachay-Balkar	(1) <b>j-</b> and <b>ch-</b> ; (2) <b>z-</b> and <b>ts-</b>	<p>There are two different dialects in Karachay-Balkar.</p> <p>No signs of /y-/ even in marginal dialects is reported.</p>
Early Kypchak	<b>y-</b>	<p>Attested as /y-/ in the Armenian and Mamluk sources.</p>
Yughur	<b>y-</b> , sometimes <b>tsh'-</b>	<p>There are a few reports from Tenishev about /tsh'-/, as if in Mandarin, but mostly /y-/ (which could be either an allophonic distribution or an unknown dialect of Yugur)</p>
Salar	<b>y-</b> , sometimes <b>dzh'-</b>	<p>Just as in Yugur, Poppe mentions a few words from Potanin's materials, where /y-/ is irregularly rendered as /dzh'-/ in the Russophone transcription, which roughly equivalent to the English /j-/, e.g. <i>dzhigirme</i>, <i>jigirme</i> as opposed to the usual <i>igermi</i>"twenty".</p>
Transoxanian Oghuz (c. 11th century)	<b>j-</b> and <b>y-</b>	<p>Confusingly attested as both /j-/ and /y-/ by al-Kashgari, but /j-/ is more certain.</p>

Turkmen	y- < *j-(?)	Because of the attestation of /j-/ in Transoxanian Oghuz, the accepted source of the Seljuk languages, we should deduce that /y-/ may in fact be a later development in Proto-Seljuk, for instance, due to the Karakhanid, Chagatai and Uzbek influence.
Azeri	0- < y-	A regular loss of /y-/, as in <i>ürək</i> < <i>yürek</i> "heart"
Turkish	y-	In some instances, /y-/ may even be weakened further or disappear, as in Azeri, e.g. /biliyor/ "he knows" > /bilior/ in the real pronunciation.
Orkhon Old Turkic (c. 9th century)	y- (?)	Commonly interpreted as /y-/, but no exact evidence
Karakhanid (11th c.)	y-	Clearly attested as /y-/ in al-Kashgari's work
Uzbek, Uyghur	y- < *zh-; j- (Kypchak Uzbek) j-, y- (Uyghur)	Presently, written as /y-/ probably due to the Karakhanid influence; originally, probably /zh-/ or /j-/ because of the close relatedness to the early Kazakh-Kyrgyz-Kypchak (see below). The /j-/ phoneme is found in the Kypchak dialect of Uzbek (e.g. <i>jaxshi</i> as opposed to the usual <i>yaxshi</i> "good").  Interestingly, Uyghur mostly uses /j-/ and /y-/ interchangeably, so they must be in an allophonic distribution.

This table shows that the pure /y-/ pronunciation is attested only within the following subtaxa:

(1) in the languages historically connected with the Orkhon-Karakhanid and Oghuz-Seljuk subgroups, even though there seems to exist some /y-/to-/j-/ allophonic distribution in Uyghur, some Uzbek dialects and some Oghuz dialects;

(2) partly, in Yugur and Salar, which also belong to the southern Orkhon-Karakhanid habitat and may have been contaminated by it, considering they are located along the Silk Road outposts, where migrations were a very common phenomenon.

(3) partly, in the /ya-/, /yu-/, /yo-/ syllables, in the languages descending from the late expansion of the Golden Horde, such as Kazan Tatar (but not the Kimak languages with an early separation, such as Karachay-Balkar). Nevertheless, even in Kazan Tatar, many speakers still report

an allophonic distribution of this phoneme, therefore a clear-cut /y-/ exists mostly in the written standard, produced more or less artificially after the 1920's, as well as in the recently Russified speech, rather than in older dialects or geographically marginal languages, such as North Crimean Tatar, Eastern Bashkir, etc. Moreover, we still have /jil/, not /yil/ "wind" before a high vowel even in the standard Kazan Tatar.

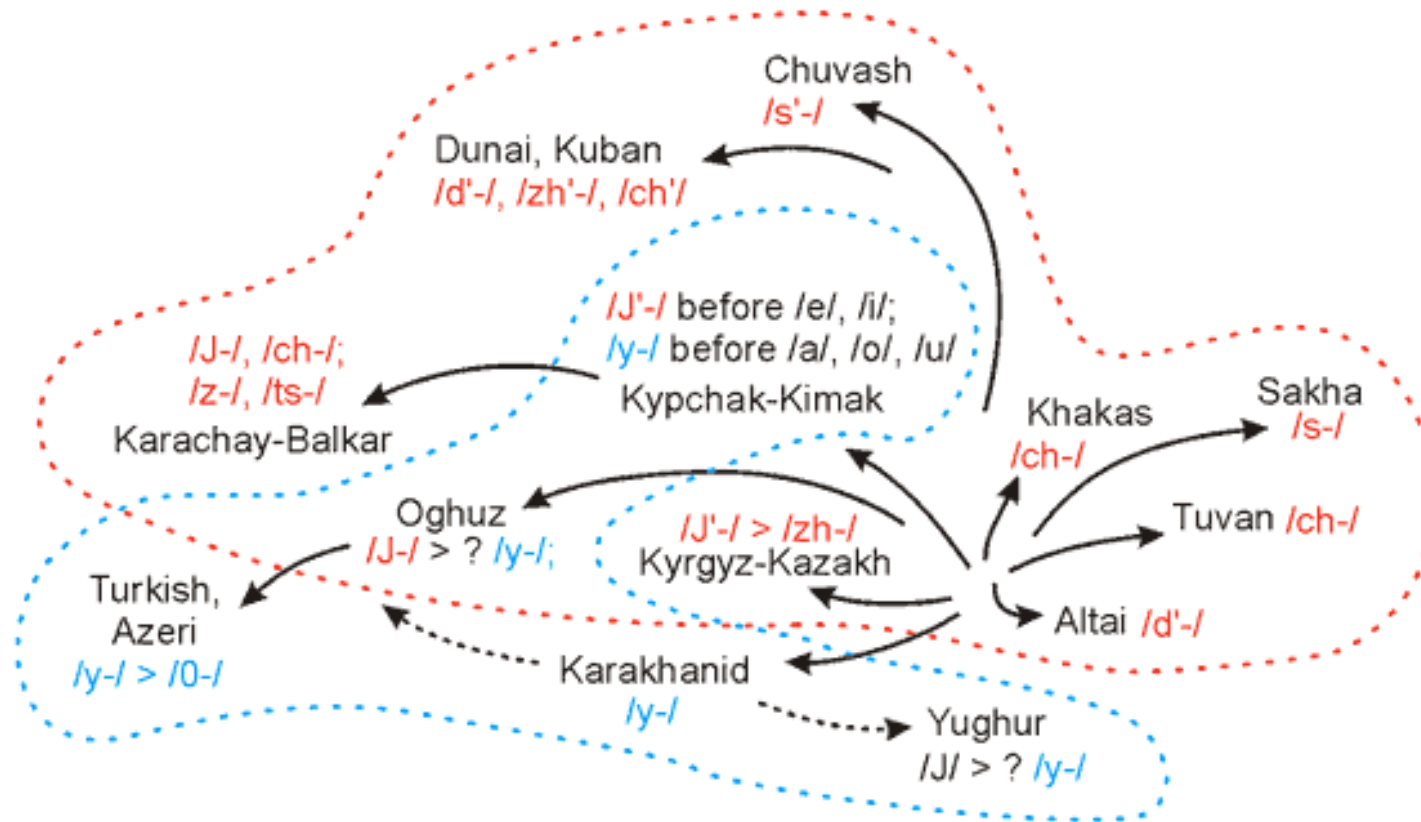
Consequently, we may conclude:

(1) Only the languages related or adjacent to the Oghuz-Orkhon-Karakhanid branch seem to have a clear-cut historical attestation of the /y-/ semi-vowel, whereas the majority of other branches with an early separation and long isolation either get jumbled data or seem to be clearly going back to something like a strongly palatalized sibilant /s'-/, /j-/, /d'-/, /ch-/ or a similar consonant sound.

This provides a purely statistical argument for our conclusion: there are more separate language branches that originally had an /s'-/- or /j-/-type phoneme than those that finally developed the /y/-phoneme. To put it in other words, it is statistically implausible that the supposed /y- / > /j-/ mutation would have occurred simultaneously and independently in so many separately existing archaic branches.

(2) As we can see in the fig. below, the distribution of the y-type phoneme seems to be located outside of the main historical diversification area of Turkic languages, therefore it appears to be a recent phonological mutation, apparently linked to the migration of the Orkhon-Karakhanid and Oghuz languages, which again implies that the development of /y/ might have been a rather unique phonological innovation in Orkhon-Karakhanid Old Turkic. This provides us with a second phono-geographical argument: only the J-type phoneme seems to be distributed near the putative homeland area of Turkic languages, not the y- semivowel.

### THE DISTRIBUTION OF THE J-TYPE vs. Y-TYPE PHONEME IN THE TURKIC LANGUAGES



As to the existence of the allophonic /y-/to-/j-/ phonological variation in the Kimak-Kypchak-Tatar languages of the Golden Horde, such as Kazan Tatar, the existence of /y-/ may be explained as *an early Oghuz influence*. As we will show below, the Golden Horde languages and Oghuz share many linguistic features at several levels, therefore this type of borrowing is well corroborated by other evidence of mutual interaction.

(3) Moreover, if /y-/ were present in the proto-form, we would rather observe phonological variations of the *semi-vowel* /y-/ (not /J-/): e.g. we would find something like /y-/ , /i-/ , /o-/ , /ê-/ , /l'-/ , /J-/ , /zh-/ in the most archaic and diversified Siberian branches *in the east* (near the historical homeland of the Turkic languages), but what we do see in that area are the phonological variations of the palatalized *consonant* /s'-/: /s'-/ , /s-/ , /h-/ , /ch'-/ , /J-/ , /zh-/ , /d'-/ , /ni-/ , /y-/. On the other hand, the expected zero phoneme resulting from the loss of /y-/ is only present in the westernmost languages, such as Azeri (e.g. *ulduz* < *yulduz* "star", *il* < *yil* "year"), and, partly, in Turkish (cf. *ılık*, but Turkmen *yili* "warm"), which marks the /y-/ -phoneme as a relatively recent and rather westernmost phenomenon connected with the spread of the Oghuz-Seljuk languages. This provides us with a *phonological diversification argument*: *if the /y-/ semi-vowel were original, there would be a range of predictable sound changes in the most early diversified branches*, but nothing of the kind is found there.

Therefore, from the evidence *internal* to the Turkic languages alone, we may conclude that the \*S- proto-phoneme in question can be placed somewhere within the range of sibilants {/s'-/ , /s-/ , /h-/ , /ch'-/ , /J-/ , /zh-/ , /d'-/}, and it could not have been similar to the /y-/ semivowel as in modern Oghuz-Seljuk languages.

Actually, this conclusion concerning the reconstruction of the Proto-Turkic \*S- is hardly novel and has been expounded several times by different authors, such as A.N. Bernshtam (1938), S.E. Malov (1952), N. A. Baskakov (1955), A.M. Scherbak (1970), as well as by the authors of the authoritative Russian publication, sometimes abbreviated as *SIGTY*, namely in its volume [*Pratyurkskiy yazyk-osnova. Kartina mira pratyurkskogo etnosa po dannym yazyka. (The Proto-Turkic language. The Worldview of the Proto-Turkic ethnicity based on the linguistic data.)*, Moscow (2006)].

**Note:** Generally speaking, *SIGTY* [*Sravnitelno-istoricheskaya grammatka tyurkskikh yazykov ("The Comparative Historical Grammar of the Turkic languages")*] is a large and verbose multi-volume Moscow comprehensive publication with detailed cross-comparative analysis of morphology, syntax, vocabulary, semiotics and other aspects of Turkic languages, produced between the 1970's and the 2000's.

As an additional quite interesting argument, the authors of *SIGTY* suggest that, since other sonants, such as \*r- and \*l-, were absent or atypical in the word-initial position, there is no reason to believe that the /\*y-/ semi-vowel, phonetically similar to a sonant, could be there either.

The opposite view, which mostly goes back to Radlov's work in the end of the 19th century is usually based on the following incorrect presumptions: (1) that the Karakhanid Old Turkic of Makhmud al-Kashgari is equal to all of the Turkic languages (in other words, that Middle Turkic = late Proto-Turkic); (2) that Orkhon Old Turkic has been correctly and uncontroversially reconstructed from the script and it reflects /y-

/, even though we hardly know the actual pronunciation in the Orkhon inscriptions; (3) that the high level of differentiation among different Turkic subgroups can be ignored, including the evidence for the maximum differences in the Siberian languages and Chuvash — in this approach the evidence from the Kimak-Kypchak-Tatar languages, for instance, may play the same role as the evidence from Sakha, and indeed this was the situation in Russian and European Turkology until the beginning of the 20th century, when most Turkic languages were officially viewed as merely dialects of each other. Even in *SIGTY*, Chuvash is still unreasonably included into the mainstream Turkic languages, at least as far as the phonological reconstructions are concerned.

As a final touch, we can describe a phonological calculation based on the above-postulated formula used in the reconstruction of the S-phoneme:

$$1/2 \text{ Proto-Chuvash } /s'-/ + 1/2 [1/3 \text{ Proto-Yakutic } /s'-/ + 1/3 (1/2 (1/2 \text{ Proto-Altay-Sayan } /ch'-/ + 1/2 (1/2 \text{ Proto-Kimak-Kypchak } /j'-/ + 1/2 \text{ Proto-Kyrgyz-Kazakh-Chagatai } /j'-/)) + 1/3 \text{ Proto-Oghuz-Orkhon-Karakhanid } /y'-/)] =$$

$$1/2 \text{ Proto-Chuvash } /s'-/ + 1/2 [1/3 \text{ Proto-Yakutic } /s'-/ + 1/3 (1/2 \text{ Proto-Altay-Sayan } /ch'-/ + 1/2 \text{ Proto-Great-Steppe } /j'-/ ) + 1/3 \text{ Proto-Oghuz-Orkhon-Karakhanid } /y'-/)] =$$

$$1/2 \text{ Proto-Chuvash } /s'-/ + 1/2 [1/3 \text{ Proto-Yakutic } /s'-/ + 1/3 \text{ Proto-Central } /ch'-/ + 1/3 \text{ Proto-Oghuz-Orkhon-Karakhanid } /y'-/]$$

It follows from this expression that *the original Proto-Bulgaro-Turkic \*S-phoneme was most likely similar to a soft palatalized /s'-/ as in modern Chuvash /s'/, Russian /sh'/ or Japanese*, hence for instance *\*s'etti* "seven" as in the Indo-European *\*septem*, not *\*yetti*, as it perhaps follows from Turkish, Azeri, Uzbek, Karakhanid and other widespread Turkic languages.

At a later stage, the phoneme began to change into a soft palatalized unvoiced */ch'/* or voiced */j'/* after the separation of Proto-Yakutic, whereas the mutation to */y'-/* was a relatively recent innovative phenomenon typical only of the southern branch of Turkic languages.

## 2. Collecting factual material

Comprehensive research in Turkology was often hindered by the large number of languages and dialects (somewhere over 50 when all the major dialects are counted) and the lack of detailed grammars and dictionaries for some of them. In many cases, the language descriptions were composed only after the 1920's or even after World War II.

As a result, most of the 19th century's Turkological classifications had originally been built upon phonological criteria alone. The grammatical features were slowly added in in the course of the 20th century, whereas detailed lexicostatistical and glottochronological analysis seems to be the thing of the recent past that appeared mostly in the 1990's.

In the present chapter, we will briefly summarize the essential *lexical, grammatical and phonological evidence collected as the basis for further examination* in the next chapters.

## **2.1 An overview of the lexicostatistical research in Turkic languages**

In the beginning of the 21st century, several authors attempted to conduct some *purely statistical* studies of the Turkic languages, in most cases without any manual analysis of grammar or vocabulary.

### **Starostin (1991)**

Sergey Starostin [STAH-res-tin] included some very detailed 110-word Swadesh-Yakhontov wordlists for 21 Turkic language in his book [*Altajskaja problema i proiskhozhdenije japonskogo jazyka (The Altaic Problem and the Origins of the Japanese language)*, Moscow (1991)]. These lists were apparently later reintegrated into the Starling database.

### **Dyachok (2001)**

A work conducted by M. Dyachok [pronounced: d-yah-CHOK] was published online as brief preliminary notes. In the introduction to his concise article, the author reminds the reader of the old geography-based classification by Samoylovich [sah-moy-LAW-vich] (1922), which had similar results, and then performs the lexicostatistical and glottochronological analysis of the 13 major Turkic languages. As a result, the Turkic languages were subdivided roughly into merely four basic subgroups (1) Bulgaric (2) Yakut, (3) Tuvan, (4) Western (= any other), which conforms to the idea that their area of maximum diversification was located somewhere in the east.

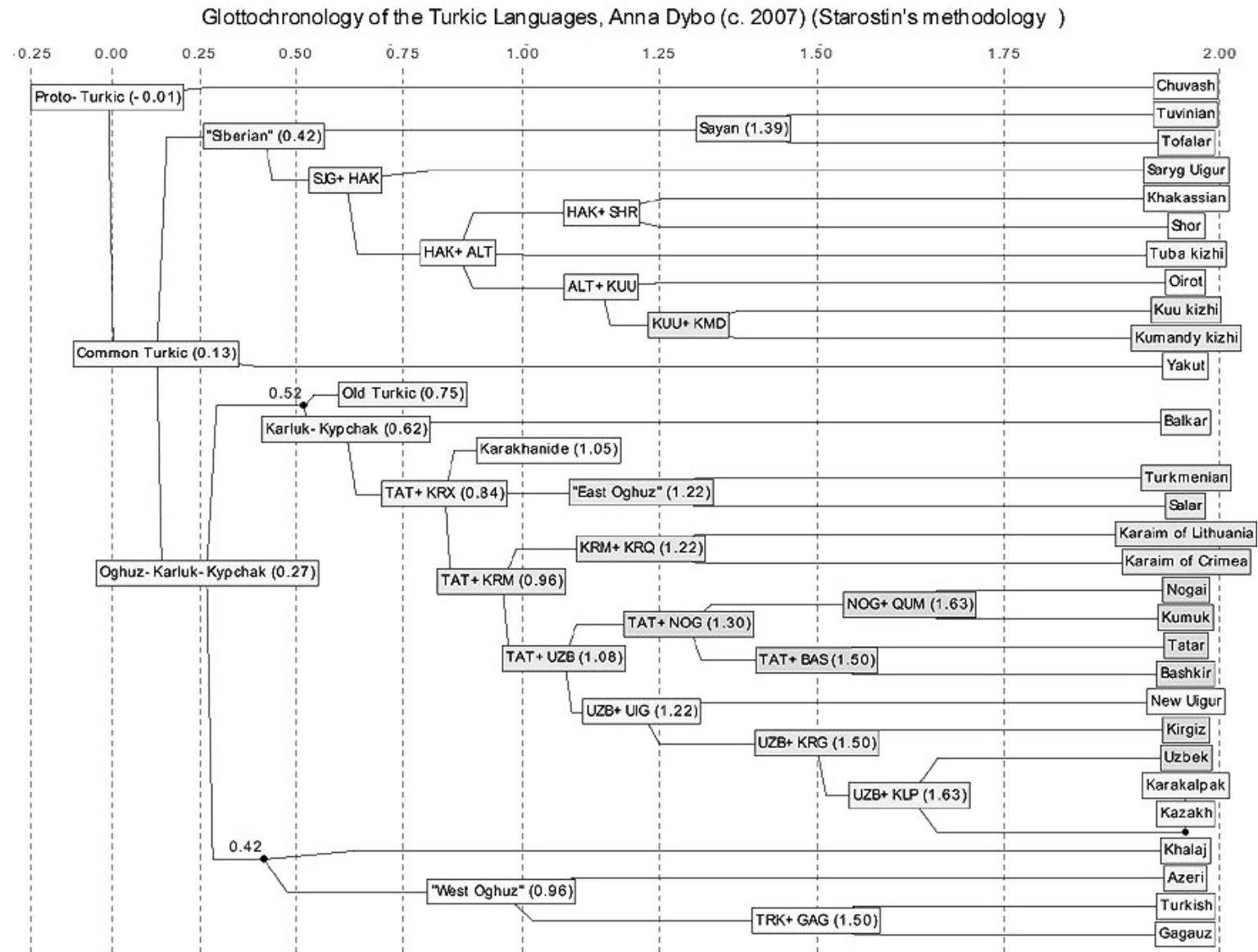
### **Dybo (2002, 2007)**

The study by Anna Dybo [AHN-nah deh-BAW] was first published in 2001 as part of the articles collected in *SIGTY* [(*Sravnitel'naja grammatika tyurkskikh jazykov (The Comparative Grammar of the Turkic languages)*]. Then, it was republished in 2007 in a separate book [Anna Dybo, *Lingvisticheskiye kontakty rannikh tyurkov. Leksicheskij fond. (The Linguistic Contacts of the Early Turks: the Lexical Fund)*, Moscow (2007)].

The study cites Dyachok as a recent lexicostatistical publication and then briefly describes its own methodology, "*All the languages, for which the 100-Swadesh wordlists could be collected from written sources, were included into our investigation. The 100-word Yakhontov-Starostin wordlists were employed, taken that they allow better accuracy [= than the classical Swadesh-100]; they were processed according to Starostin's methodology by excluding the recognizable borrowings and employing the STARLING program [...]*"

As a result, the following dendrogram was obtained:





Dybo, Anna, *The Chronology of the Turkic languages and the Linguistic Contacts of the Early Turks* (2006)

There also exists a second version of this dendrogram that drastically differs from the first one, because of some kind of unexplained procedure that was applied to synonyms. This is slightly confusing and may result in the underestimation of the dendrogram's significance, however the first tree above (with the synonyms included) partly matches the outcome obtained in other investigations. Apart from such unconventional points as (1) the splitting of Turkmen and Turkish between two different taxa, (2) the positions of Yugur and Salar, (3) the slightly misplaced Kazakh (which cannot be directly related to Uzbek) and Uzbek position (which is known historically to be related to Uyghur), it is in fact in relatively good correspondence with other studies. However, the glottochronological part based on Starostin's formulas should be taken with a grain of salt.

It should also be noted that *the use of shorter 110-word lists results in lower statistical robustness* than in the current series of publications that uses larger 215-word lists. Nevertheless, this work has an advantage of representing a greater set of languages, especially those of the Altay-Sayan area, which are normally underestimated or omitted in other studies.

### **ASJP (2009)**

Another example of a *phonostatistical* research that merits mentioning is the automated dendrogram built by the Automated Similarity Judgment Program for most languages of the world. Here's a preliminary and simplified first-approximation *phonostatistical dendrogram of Turkic languages* (gif) from 04/2009.

The study was based on a simple 40-word list. Many branches seem to be mispositioned, apparently due to certain limitations of the ASJP's initial approach, however you can see the early separation of Proto-Chuvash, then Proto-Oghuz, and then the rest of the languages, which is partly consistent with the conclusions obtained in the present work and other studies.

### **Herein (2009, 2012)**

To prepare a lexicostatistical research for this publication, it was decided to use the readily available 200-word Swadesh lists from Wiktionary.org.

After verifying and correcting the available materials, building some new lists for absent languages (such as Khakas, Tuvan, Altai) (2009), composing a php-program to do all the routine calculations, performing some additional meticulous examinations and adding some new lexical material thus expanding the lists to 215 entries (2012), another lexicostatistical study named *The Lexicostatistics and Glottochronology of the Turkic languages* was finally produced.

It should be noted that the lexicostatistical figures obtained in 2009 and 2012 sometimes differed significantly from each other, because of different approaches used to account for the unavoidable synonymy. The 2009 approach had been much too basic and consequently was significantly enhanced in 2011-12, which included both reexamining the original lists and introducing changes into the program application, so the present version is to be considered more correct.

Most borrowings (Persian, Arabic, Mongolian, Russian, etc) were excluded wherever possible, so only the verified cognates were counted in the final glottochronological section of the study. In the doubtful cases the cognacy was determined according to the [*Etymologicheskij slovar chuvashskega jazyka (The etymological Dictionary of Chuvash)*, by M. Fedotov; volume 1-2, Cheboksary (1996)] and sometimes using the [*Etymologicheskij slovar tyurkskikh jazykov (The etymological Dictionary of the Turkic languages)*, E. V. Sevortyan, Vol. 1-7, Moscow (1974-2003)].

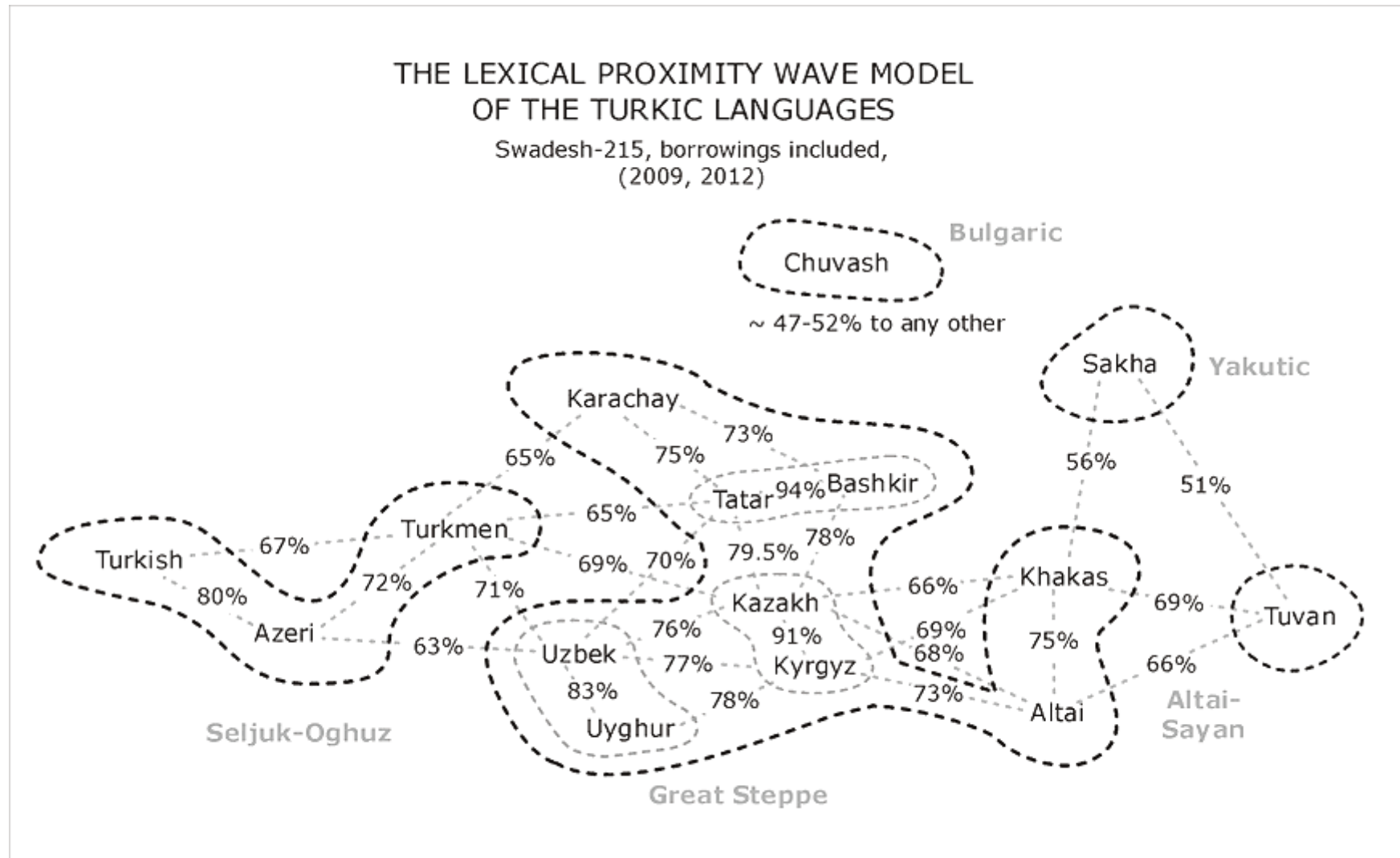
The lexical lists presently differ from the Wiktionary.org materials and are available online as a Word document

As the final outcome of the study, several lexicostatistical matrices of Turkic languages were built.

The Lexicostatistical Matrix of Turkic languages, Swadesh-215 (02.2012), borrowings excluded														
	Chuvash	Sakha	Tuvan	Khakas	Standard Altay	Kyrgyz	Kazakh	Uzbek	Uyghur	Karachay	Bashkir	Tatar	Turkmen	Azeri
Sakha	51.9%													
Tuvan	49.3%	57%												
Khakas	52.8%	61.3%	71.9%											
Standard Altay	50.9%	55.9%	69.3%	75.6%										
Kyrgyz	57.9%	59.6%	63.3%	70.3%	74.6%									
Kazakh	58.2%	59.4%	61.6%	68.1%	69.9%	92%								
Uzbek	61.1%	57.8%	58.2%	65.3%	66.3%	82.9%	82.8%							
Uyghur	59.2%	59%	61.7%	65.7%	70.2%	83.8%	81.9%	86.3%						
Karachay	57.5%	60.8%	58.7%	65.1%	65.2%	77.8%	78.3%	74.6%	77.1%					
Bashkir	58.3%	59.4%	59.9%	67.1%	69%	82%	79.9%	76.1%	78.5%	77.4%				
Tatar	59.4%	60.7%	60.2%	68.2%	70.1%	83.9%	82.1%	78%	79.6%	79.2%	94.9%			
Turkmen	55.6%	55%	54.7%	61.2%	59.5%	71.2%	71.9%	75.9%	71.7%	69.2%	71.9%	69.8%		
Azeri	55.6%	51.8%	51.8%	56.4%	58.4%	66.9%	67.8%	70%	68.8%	66.9%	66%	68.4%	78.2%	
Turkish	54.9%	52%	50%	53.8%	54.4%	64.9%	64.8%	67.2%	66.7%	64.2%	62.8%	65.6%	73.6%	86%

Considering that *an accurate analysis is supposed to include phonological, grammatical, historical and other non-lexical evidence*, the lexicostatistical data alone are most likely insufficient to build a complete dendrogram of the Turkic languages at this point,

However, we can use the values in the table to build *a wave model of Turkic languages that would reflect the mutual language intelligibility* through the calculated relationships in the basic vocabulary. The wave model should be based on the *borrowings-included* matrix, because it is supposed to represent the mutual intelligibility as it is, without any exclusions, for this reason you may notice some small discrepancy in percentages with the table above.



The wave model of the Turkic languages with borrowings included from

[*The Lexicostatistics and Glottochronology of the Turkic languages* (2009-2012)]

## 2.2 Dissimilar basic lexemes in the Turkic languages

Another brief lexical table prepared in 2009 included a visual overview of certain lexemes that are known to be *dissimilar* within the core Turkic languages. These lexical data help to pick up dissimilarities between otherwise closely related groups and assist in identifying large supertaxa.

Dissimilar Basic Words in the Turkic languages									
<p><b>Red</b> is a more ancient layer associated with the Siberian Turkic languages, <b>brown</b> marks the Oghuz-Turkmen innovations; <b>blue</b> is a more recent layer probably connected with the spread of the Gökturks; <b>green</b> marks probable "Central Turkic" innovations; <b>orange</b> marks the Altay-Sayan (Tuvan + Khakas + Altai) innovations; <b>purple</b> marks the Yakutic innovations or otherwise differentiated Yakutic words; gray and black are "other" or unclassified. Borrowings may be included.</p>									
	Turkish Azeri	Turkmen	Uzbek Uyghur Karakhanid	Kazan Tatar Karachay	Kazakh	Kyrgyz	Khakas	Tuvan	Sakha
	Seljuk	Oghuz	Karkhanid-Chagatai	Kimak-Kypchak	Kazakh	Kyrgyz	Yenisei-Kyrgyz		Yakutic
not (adj, nouns)	Tk. deGil; Az. deyil	däl	Uz. emas; Uy. emes Kh. ermes	KT. tügel; KB. tüyse	emes	emes	Kh. nimes; chox Al. emes; d'ok	eves; chok	suox
here	Tk. burada; Az. burada < *bu ara-da	shu tayda; bäri	Uz. buyerda; Uy. buyerde; mana K. munda	KT. monda, bireda; KB. mında, blaida	mûnda	mında	Kh. mında Al. mında	mında	manna
there	Tk. orada; Az. orada < *o ara-da	o tayda; ol yerde	Uz. uyerda; Uy. uyerde;	KT. anda, shul zherde; KB. anda, alaida	onda	anda	Kh. anda Al. anda	aNa:	onno

how	Tk. nasil; Az. nechê	nähili	Uz. qanday Uy. qandaq	KT. nichek; KB. qalay	qalay	qanday	Kh. xaidi Al. kandiy	qandig	xaidax
many	Tk. chok; Az. chox	köp	Uz. kûp Uy. köp Kh. talim; kûp	KT. kûp KB. köp	köp	köp	Kh. köp Al. köp	xöy	elbex, ügüs
wide	Tk. genish; Az. genish	giNish; giN	Uz. keN Uy. keN Kh. keN	KT. kiN KB. keN	keN	keN	Kh. chalbax Al. d'albak	kalbak, chalbak	kieN
forest	Tk. orman; Az. orman	tokay	Uz. ûrmon Uy. ormanliq	KT. urman B. aGach	toGay; orman	tokoy; orman	Kh. agas; Al. arka	arga, arig	tïa
root	Tk. kök; Az. kök	kök	Uz. ildiz Uy. iltiz Kh. yildiz	KT. tamir KB. tamir	tamir	tamir	Kh. tazil; chilige Al. tazil	t.azil	silis
bark (n)	Tk. kabuk; Az. qabiq	gabik	Uz. qobuq Uy. qovzaq	KT. kabik KB. qabuq	qabiq	qabiq	Kh. xabix Al. chobra	chövure:	xatiriq
flower	Tk. gül "rose"; chichek Az. gül; chichê k	gül	Uz. gül; chichak Uy. gül; chichek Kh. chichek	KT. göl; chêchêk KB. gül; gokka	gül; shêshêk	gül	Kh. chaxayax Al. chechek	chechek	sibekki
fat (n)	Tk. yaG; Az. yaG	yaG	Uz. yoG; may	KT. may;	may	may	Kh. üs, zhaG	üs, chaG	sïa



	Az. yaG;		Uy. yaG; may	KB. jau			Al. üs		
nose	Tk. burun; Az. burun	burun	Uz. burun; Uy. burun	KT. borin; KB. burun	mûrin	murun	Kh. purun, tumzux; Al. tumchuq	t.umchuq	murun
hand	Tk. el; Az. êl	el	Uz. qûl Uy. qol Kh. elig	KT. kul; KB. qol	qol	qol	Kh. xol Al. qol	xol	ili:
liver	Tk. (kara)chiGer Az. chiyer	bagir	Uz. zhigar; baGir; Uy. jiger; beGir Kh. baGir	KT. bawir; KB. baur	bawir	boor	Kh. paar Al. buur	p.aar	bïar
think	Tk. düşhün- Az. düşhün-	öyt-	Uz. ûyla-; Uy. oyli-	KT. uyla-; KB. oymla-	oyl-	oyl-	Kh. sagin- Al. sanan	p.od-	sana:
live	Tk. yasha- Az. yasha-	yasha-	Uz. yasha-; Uy. yashi-	KT. yashê-; KB. jasha-	zhas-	zhash-	Kh. churt- Al. d'ür-	churtt-	olor; sirit
say	Tk. de- Az. de-	diy	Uz. ayt-; de- Uy. eyt-; de- Kh. ay-; de-	KT. êyt- KB. ayt-	ait-; de-	ait-; desh	Kh.cho:xt- Al. ayt-	chug-; t.e:-	die, et
sky	Tk. gök Az. göy	gök	Uz. kük; asman Uy. kök; asman	KT. kük KB. kük	kök (rare);aspa n	kök (rare);asm an	Kh. tigr Al. teNeri	t.e:r	xalla:n
burn (intr.)	Tk. yan-	öt-; yan-	Uz. yon-	KT. yan-	zhan-	köy-; zhan-	Kh. köy-	kiv-	umai

	Az. yan-		Uy. yan-; köy-	KB. jan-			Al. küy-		
night	Tk. geche Az. geche	gije	Uz. tün Uy. tün Kh. tün; kecha	KT. tün KB. köche	tün	tün	Kh. tün Al. tün	t.ün	tü:n
yesterday	Tk. dün Az. dünên	düyn	Uz. kech Uy. tünügün	KT. kichê KB. tünene	keshe	keche	Kh. kiche Al. keche	t.ü:n	beHehe:
evening	Tk. aksham Az. axsham	agsham	Uz. okshom; kecha Uy. axsham; keche; Kh. axsham	KT. kich KB. ingir	kesh	kech	Kh.i:r Al. engir	kezhe:	kiehe
big	Tk. büyük Az. böyük	uli; chishik	Uz. büyük; katta Uy. büyük;yoGan,zor; r;chong Kh. uluG	KT. zur KB. ullu	tülken; zor	chong	Kh. ulug; Al. d'a:n	ulug	ulaxan
child	Tk. choJuk Az. ushaq, chaga	chaga	Uz. bola; Uy. bala	KT. bala; sabii KB. sabii	bala	bala	Kh. pala; Al. bala	urug	oGo
face	Tk. yüz; Az. üz	yüz	Uz. yuz Uy. yüz	KT. bit; yöz; KB. bet	bet; zhüz; shıray	bet	Kh. sıray; Al. d'üs; chıray	shıray	sirey

island	Tk. <i>ada</i> ; Az. <i>ada</i>	<i>ada</i>	Uz. <i>orol</i> ; Uy. <i>aral</i> ; Kh. <i>utruG</i>	KT. <i>utrau</i> ; KB. <i>ayrimkan</i>	<i>aral</i>	<i>aral</i>	Kh. <i>oltirix</i> ; Al. <i>ortolik</i>	<i>ortuluk</i>	<i>arī</i>
owl	Tk. <i>baykush</i> Az. <i>baykush</i>	<i>baygush</i>	Uz. <i>boygoli</i> ; Uy. <i>baykux</i>	KT. <i>yabalaq; ökö</i> (di al.) KB. <i>uku</i>	<i>üki</i>	<i>ükü</i>	Kh. <i>tasxa</i> ; Al. <i>mechirtke</i>	<i>mezhergen</i>	<i>mekchirge</i>
tomorrow	Tk. <i>yarin</i> Az. <i>sabah</i>	<i>ertir</i>	Uz. <i>ertaga</i> Uy. <i>ete</i>	KT. <i>irtêgê</i> ; KB. <i>tambla</i>	<i>erteN</i>	<i>erteN</i>	Kh. <i>taNda</i> ; Al. <i>erten</i>	<i>erten; t.a:rtā</i>	<i>sarsin</i>
voice	Tk. <i>ses</i> Az. <i>sês</i>	<i>ses</i>	Uz. <i>ovoz</i> Uy. <i>awaz</i>	KT. <i>tavish, avaz</i> KB. <i>auaz</i>	<i>dawis</i>	<i>ün</i>	Kh. <i>ün</i> Al. <i>ün</i>	<i>ün</i>	<i>kuolas, saNa</i>
wet	Tk. <i>yash</i> Az. <i>yash</i>	<i>öl</i>	Uz. <i>ho'l</i> Uy. <i>höl</i>	KT. <i>yuesh, dïmlï</i> KB. <i>mïli, Jibigen</i>	<i>dïmqil</i> ; <i>su</i>	<i>nïm, nïmdu:</i>	Kh. <i>öl</i> Al. <i>ülüş, chïqt</i> <i>u</i>	<i>öl, mö:n, shal</i>	<i>incheGey, u:la:x,</i> <i>si:kte:x</i>

### 2.3 The comparison of phonological and grammatical features

#### Mudrak (2002, 2009)

The multivolume Moscow edition *SIGTY. Regionalnyiye rekonstruktsii* ("The Comparative Grammar of Turkic languages. Regional Reconstructions.") (2002) included an abbreviated article by Russian Turkologist Oleg Mudrak [aw-LEG moo-DRAHK; the name is etymologically akin to *mudryj* "wise, sagacious"] *Ob utochnenii klassifikatsii tyurkskikh yazykov s pomosh'yu morfologicheskoy lingvostatistiki* (On the clarification of the Turkic languages classification by means of morphological linguostatistiki). It was subsequently republished in full as a separate book in 2009, and then briefly reviewed in a public lecture on *the history of Turkic languages* (available at youtube.com and as an magazine article).

The study uses *a unique statistical analysis of 96 morphological and phonological features* counted up for as many as 42 Turkic languages and major dialects, and builds up trees with glottochronological dates (though based again on the apparently incorrect Starostin's glottochronological formulas), checking them for consistency with the major historical events. This purely morphostatistical analysis is an extremely interesting and apparently completely novel approach in historical linguistics. The obtained dendrograms roughly coincided with the present study by about 80%, though differed in certain aspects.

\* \* \*

The purely grammatical approach by Mudrak prompted us to take a closer look at the morphological features, which are well-known to be more resistant to borrowings than common words thus providing more robust results. Finally, a similar study of phono-morphological differences within the Turkic languages was conducted (2009).

The following table contains a list of *certain phonological and grammatical features known to be different across Turkic languages*, so studying them helps to establish the exact order of their taxonomic diversification.

It should be acknowledged that the former analysis of phono-morphological features by Mudrak (2009) seems to be more detailed, particularly as far as the number of included languages is concerned. However, even though many additional grammatical and phonological characteristics are not explicitly mentioned in the table of phonological and morphological differences, they are often described below under paragraphs for specific Turkic languages.

Much of the morphological and phonological data in the table have been collected from the encyclopedic edition [*Jazyki mira: Tyurkskije jazyki (The Languages of the World: The Turkic Languages)*; editorial board: E. Tenishev, E. Potselujevskij, I. Kormushin, A. Kibrik, et al; The Russian Academy of Sciences (1996)], which is a detailed, comprehensive and authoritative publication consisting of articles by specific authors and brief phonetical and grammatical descriptions of each Turkic language. Other data were collected directly from grammar books on specific languages.

Some of the phonological and morphological differences within the Turkic languages

The table may contain simplifications in transcribing vocal harmony

	y-/ J-	-G-/ -w-	-d-/ -y-	b-/p- t-/d- g-/k- G/q-	Instrum ental case	Other cases	Plur al	Dativ e	"Perfect" Participle	Negation of adjectives , nouns	"We did" ending	"We do" Aorist ending	"I do" Aorist ending	Use of <i>tur-</i> or any other copula	Future Tense	someone, somewher e, no one, nowhere	you (plural)
<b>Chuvash</b>	s'-	-v-	-r-	p-, t-, k-, x-	-pa, -pe	Goal- directed -shan, -shen	-sem	-a, -e	—	mar	-r-âmâr, -r-êmêr	-âpâr -êpêr	-âm -êm	—	-at-, -et- -0-	ta-kam; tashta; nikam ta; nishta ta	esir
<b>Sakha</b>	s-	-0:-	-t-	b-, t-, k-, k-	-nan	Partial -ta; Compar. -ta:Gar;	-lar, - ler, - lor, - lör, - nar, - ner, -dar, -tar, etc	-ga	-bit, -bît	suox; buol- batax	-ti-bît/bit, -li-bît/bit	- bît/bit, -pît/pit	- bîn/bin, -pîn/pin	verb-an+ tur + pronoun = past tense	-ïah-; -a:ya- / -eye-i = optative (apprehe n- sive)	kim ere, xanna ere, kim da + negative, xanna da + negat.	ehigi

<b>Tuvan</b>	ch-	-0:-	-d-	weak semivoiced : strong unvoiced: *q > x	Directiv e -dīva, -dive,- duva, -düve,- tīva, etc	-lar, - ler, - nar, - ner, - tar, - ter, - dar, - der	ga/ge, -gan, etc -ka/ke	eves; chok	-dī-vīs	-vīs, - vis - vūs, - vus	men	verb + p + tur (chīdīr, olur) + pronoun =Present	-īr-; Gai/gei, qai/kei = optative	bir-(le) kizhi; bir-(le) cherde; kīm-da: + negativ; kaida-da: + negative	siler	
<b>Tofalar</b>	ch-	-0:-	-d-	weak semivoiced : strong unvoiced d	Partial -da, -de, -ta, -te	-lar, - ler, - nar, - ner, - tar, - ter	Ga/Ge -Gan/Gen, -qan/ qen -qa/qe	emes	-dī-vīs	-bis	men	verb + p + turu (chīētīrī, oluru) + pronoun = Present tense	-ar/er/īr/īr -; Gai/gei, qai/kei = optative	-- qum-ta: + negat. --	siler	
<b>Khakas</b>	ch-, n'-	-0:-	-z-	p-, t-, k-, x-	-naN, -neN	Directiv e -lar, - ler, -nar, - ner, - tar, - ter	-ga/ge - xa/ke, - na/ne, -a/e	-Gan/gen, -xan/ken	nimes; chox	-dī-bīs	-bīs/bīs -pīn/pīn - mīn/mī n; - mīs/mī s	-bīn/bīn -pīn/pīn - mīn/mī n; - mīs/mī am	verb + (p) + tur + pronoun = Audative or Archaic past;	-ar/er/r-; Gai/gei, qai/kei = optative	kem-de, xayda-da; kem-de + negat. xayda-da + negat.	sirer

<p><b>Kumandy</b></p>	<p>ch-, n'-</p>	<p>-0:-</p>	<p>-y-</p>	<p>b/p-, t-, k-, k(q)-</p>		<p>Directiv e -za, -ze, -sa, -se</p>	<p>-lar, - ler, -nar, -ner, -dar, der, -tar, - ter,</p>	<p>-ga, - ge, - ka, - ke -a, -e, etc</p>	<p>-gan, - gen, -kan, -ken</p>	<p>eves, emes; chok, chox</p>	<p>-di-bis, -di-bis, - di-vis</p>	<p>-bis, - bis, -pis, - pis</p>	<p>-im, - am</p>	<p>verb + ip + tur + pronoun = Audative past; verb + a/e + tur + -ar + pers ending = Present Future;</p>	<p>-ar/er/r-; -ad, -ed Gai/gei, qai/kei = Optative</p>	<p>kem-de, kayda-da; ---</p>	<p>sner, snir</p>
<p><b>Standard Altai</b></p>	<p>d'-</p>	<p>-0:-</p>	<p>-y-</p>	<p>b-, t-, k-, q-</p>			<p>-lar, - ler, - lor, - lör, -dar, der, -dor, dör, -tar, - ter, -tor, - tör</p>	<p>-ga, - ge, -go, - gö, etc</p>	<p>-gan/gên, -kan/kên</p>	<p>emes; d'ok s, is/is, - ik/ik</p>	<p>- (i)bis/(i)bi -bis, -bis,</p>	<p>-bin/bin -pin/pin -min/min</p>	<p>verb + dir + pers ending = audative past; verb + a/e + dir + pers ending = Present Continuou s; verb + ip/ip + tur + d + pers ending = Past Continuou s;</p>	<p>-ar/er/r-; -at/et-; Gai/gei, qai/kei = Optative</p>	<p>kem-de, *kayda- da; ---</p>	<p>slerler</p>	

<b>Kyrgyz</b>	J-	-0:-	-y-	b-, t-, k-, q-	-	-	-lar, - ler, - lor, - lör,  -dar, der,  -dor, dör,  -tar, - ter,  -tor, - tör	-ga, - ge, - go, - gö, - ka, etc	-gan-	emes	-dik, etc	-(i)biz	-mîn	verb + iptir = audative past;  verb + ip + tur (otur, Jat, Jur) + pronoun = Present Continuos;	-ar;  Gai/gei, qai/kei = Optative	(kimdir) birö;  kayda-dir (bir Jerde);  ech kim; ech kaida, ech Jerde	siler, sizder  siz (polite)
<b>Kazakh</b>	J-, zh-	-w-	-y-	b-, t-, k-, q-	-men, - pen	-	-lar, - ler,  -dar, der,  -tar, - ter,	-Ga, - ge, -qa, - qe	-Gan, - Gen  -qan, -qen	emes	-diq, -dik	-miz, - miz	-bîn/bin -pîn/pin  -mîn/mi n	verb + ip + tûr (otur, Jatirt, Jür) + pronoun = Present Continuos;	-ar/er/r;  -baq/bek- , -paq/pek- , - maq/mek -	êlde-bireu, êldekim  bir Jerde  esh kim; esh kaida, esh Jerde	sender; siz, sizder(polit e)



<b>Uzbek</b>	y-	-G-	-y-	b-, t-, k-, q-	-	-	-lar	-ga	-gan, - qan, -miş-	emas	-dik; - dimiz <i>(dialectical variation)</i>	-(i)miz	-man	verb + ip + tûr (ûtir, yot, yür) + pronoun = Present Continuos;	-a-, -y-; -ar/r;	allakim, kimdir -- hech kim; hech qayerda;	siz
<b>Uyghur</b>	y-	-G-	-y-	b-, t-, k-, q-	-	-	-lar, - lêr	-gê, - qa, - ka, - kê, -qê	-Gan	êmês	-duk, -tuq	-(i)miz	-mên	verb + ip + tur (oltur, yat, yür) + pronoun = Present Continuos;	-i--; -ar/r;	kimdu, biri -- hech qaysi , hech kim; hech yerde;	silêr, siz (polite)
<b>Chagatai</b>	y-	-G-	-y-	b-, t-, k-, q-	-	-	-lAr	-Ga, - gâ, -qa, - kâ	-Gan, - Gân - miş- (rare)	e(r)mäs, y oq	-dîq (or similar)	-(i)bîz	-men (-Am)	noun + dur(ur); verb + -A + dur- pronoun; verb + Yp + -dur;	-Gu-	kishi, -- -- -- --	siz, sizlär

<b>Baraba</b>	y-	-y-	b-, t-, k-, q-	-	-	-lar, - nar, - tar	-qa	-Gan	tügil	-dıq, etc	-bıs,	-mın, -Am	verb + ip + tur (otır, yat) + pronoun = Present Continuos (rare);	-İR; -		silär; siz (polite)
<b>Karachay</b>	J-, ch-	-w-	-y-	b-, t-, k-, q-	-	-la, - lê	-ga/- xa/ ge, - na/ ne, - a/e	-Gan/gen	tüyül	-dıq, -duk, -dük, etc	-bız, - biz, etc	-ma, - me	verb + a/e + tur + pronouns = Present Continuou s;	-İR; - rıq/nıq/lı q;	kim ese da, qaida ese da, --	siz
<b>Tatar</b>	y- Ji- Je-	-w-	-y-	b-, t-, k-, q-	Comparat. -day, - tay, -dêy, - dıy, etc. Locat- Temp. -dagı, - tagı, -dêge	-lar, - lêr, - nar, - nêr	-ga, - gê, - ka, - kê; - na/nê, -a/ê	-gan, -kên	tügel; participle + pers. ending + yuk	-dik, etc	-bız, etc	-m(în)	noun (3rd pers) + - dYr, -tYr	-İR; -achak;	kemder; kaidadır; berkaida; ber kem (dê), hichkem; (ber) kaida da hich ber Jirdê;	sez

Cuman-Polovtsian			-y-	b-, t-, k-, q-			-lar, - ler	-Ga, - ge, - qa, - ke; -a, -ê	-mYsh-			-büz	-man, -men	noun (3rd pers) + - dYr, -tYr	-Gai/- gei, -kai/-kei		siz
Turkmen	y-	-G-	-y-	b-, d-, g-, G-			-lar, - ler	-a, -ä, -e; -na, - ne	-mYsh Used only as audative particle	dêl, participle + pers. ending + - ok		-dYk	-Ys -în, -in, -un, -ün	verb + ip + dur (otîr, yat) + pronoun = Present Continuos; verb + ip + tîr + pronoun = Past Audative; verb, noun (3rd pers) + -dYr, - tYr	-ar, -îr; -Jak, - Jek (no endings)		siz
Azeri	y-	-G-	-y-	b-, d-, g-, G-			-lar, - ler	-a, -ê	-mYsh- Used as audative particle and perfect tense	deyil		-dYg	-Yg -êm; -am	verb, noun (3rd pers) + -dYr, - tYr	- (y)acak( G- - (y)ecek( G-)	hech kim	siz

<b>Turkish</b>	y-	-G-	-y-	b-, d-, g-, G-	-	-	-lar, - ler	-(y)a, -(y)e	-mYsh-  Used as audativ e particle and perfect tense	deil, de(G)il	-dYk	-Yz	-im, -im, -um, -üm	verb, noun (3rd pers) + -dYr, - tYr	-ar, -ır; -acak(G- ), -ecek(G- )	kimse, bir shey; hich kimse, hich bir shey	siz
<b>Khalaj</b>	y-	-G-	-d-	b-, t-, k-, q-	-la	Locative -cha	-lar	-ka, - qa, - yâ	-mYsh-	daG	-dimiz, -dYk < Azeri	- (i)miz, -uq < Azeri	-Vm	är (conjugate d copula)	-(i)Ga		siz
<b>Karakhanid</b>	y-	-G-	-ö-	b-, t-, k-, q-	-in, -in, - un, -ün, - nin, -nin	-	-lar, - lär	-qa, - kê, -Ga, - gê, -a, -ê, -Garu, -gerü	-miş-, - mish; -Gan-, -gen-, - qan, -ken-	ärmês; yok	-dimiz, -duk	-biz, - miz		ol (3rd pers. copula)	-Gay, - gey, - qay, -kêy		siz

<b>Khorezmian</b>	y-		b-, t-, k-, q-	-n, -in, - in, -un, - ün, -an, - än		-lar	-qa, - kä, -a, -ä	-miş-, -mish-	ärmäz, ärmäs;  däGül, dügül (rare); yok	-duq, -diq	-biz	-män	er-;  -b turur = perfect past;  -a turur = repetitive present	-Gay, - gäy, - qay, - käy, -Ga, -gä, -qa, -kä		(siz)
<b>Old Uyghur (Kojo)</b>	y-	-	-ö-, -d-, -z-, b-, t-, k-, q-	-in, -in, - un, -ün, - nün, -nin	Equative -cha	-lar, - lär	-qa, - kä,  -Ga, - gê, - Na, Nä;  -Garu, -gärü	-miş-, -mish-	täGül;  ärmäz	-timiz,  -dimiz	-biz, - miz, - biz - miz	-män	ärür (copula)	-Gay, - gäy  -Galir;  -tachï, - dachï		siz
<b>Orkhon Old Turkic</b>	y-?	-G-, -G	-ö- b-, t-, k-, q-	-in, -in	Equative -cha	-lar, - lär	-qa, - gä,  -ya, - yä;  -Garu, -gärü	-miş-, - mish;  -Gan-	-; jok	-timiz,  -dimiz	-biz	-män	er-	-tachï, - dachï		siz

Salar	y-	-G-	-t-, -y-	weak semivoi ced : strong unvoice d			-lar, - lär, - ner	-Ga, - ge, -Gan, - gen; -qa, - ke, -a, -e	-Gan, - gen; -mish-	emes, emes-tür, emes-ar, yox-tür				noun + dır (idir-, oN;irar); adj + dır (idir + oN;irar); verb + p + o(r) + (tur) = Present I; verb + qu(r) + (tur) = Future I; verb + q/Gan + dır = Past II;	- ar/er/ir/ir ; -qur/Gur --	k'em-ter -- niNgi --	seler
-------	----	-----	-------------	---	--	--	--------------------------	---	---------------------------	---	--	--	--	---	---	-------------------------------	-------

Yugur	y- tsh-,	-G-	-d-	weak semivoiced : strong unvoiced	Compar.	-lar, - ler, - nar, - ner,	-Ga, - ge, -qa, - qe	-Gan	emes-tro; yoqer, yok-tro, yoq-pe-tro					i:re = copula;	verb + Gan + tır = Present Tense;	-ar; -qış-tro, -Gış-tro	qım-er, nıer	siller
						-daG, - deg, -taG, -teg								-dar, -der, -tar, - ter		verb + qışh + tro = Future;		
														verb + Gan + tro = Past II;	-Gi, -ge, -kī, -ke			
														verb + ip/ip + tro = Past III;	-qır/Gır			

### 3. Making Taxonomic Conclusions

With all the lexical and grammatical material collected in the previous chapter, we can finally get down to the analysis of each Turkic branch. Then, we will be able to attempt to make taxonomic conclusions concerning the position of each language in the phylogenetic dendrogram.

**Note:** *Taxon* is a general concept of classification science borrowed from biology which encompasses other subdivisions, such as *group*, *family*, *macrofamily*, etc. However for all practical purposes, we do not usually distinguish between (*sub*)*group* and (*sub*)*taxon* in this article. The usage of expression "*the (Name) taxon*" is thought to be equivalent to "*the (Name) languages*". The term "*family*" cannot be used except for the language taxa of high order with a temporal separation of more than 5000 years, e.g. "*the Indo-European family*", but hardly "*the Turkic family*", except maybe in the context where it would be necessary to underline the early separation of Proto-Bulgaro-Turkic from Proto-Altaiic.

#### The Bulgaric subgroup

Chuvash, the only modern-day representative of Volga Bulgaric within the Bulgaric taxon, was definitively shown to be related to Turkic by Nicholas Poppe [*Chuvashskij jazyk i jego otnoshenije k mongolskomu i tyurkskim jazykam (Chuvash and its relatedness to Mongolian and the Turkic languages)*, Nicholas Poppe (1924)]. Poppe established regular phonological correspondences between Chuvash and other Turkic languages. In his work, he listed several influential Turkologists (Adelung (1820), Rask (1834), Ramstedt (1922-23)) who had understood and accepted the Turkic origins of Chuvash long before his publication. Moreover, according to Alexander Samoylovich, Poppe had shown that "*the Chuvash and Bulgaric languages do not stem from "Proto-Turkish" (z-group), but rather from the common progenitor of both of these groups*", thus setting Chuvash aside from the rest of the Turkic languages. [Alexander Samoylovich, *K voprosu o klassifikatsiji turetskikh jazykov (Towards the question of the classification of Turkish languages // The Bulletin of the 1st Turkological Congress of the Soviet Union (1926); reprinted in the collection of his works (2005)*].

This positioning of Chuvash within the Turkic tree has changed little ever since. For this reason, Chuvash has not been considered herein in much detail, mostly because of its evidently early separation that does not cause much controversy among scholars.

#### Some of the exclusive Bulgaric features

##### Bulgaric phonology



(1) The famous Bulgaric *rhotacism* vs. the Turkic Proper *zetacism*, or the persistent use of /-r/ where other Turkic languages normally have /-z/ (though in some cases -r- can also be found in certain positions in Turkic Proper as well, for instance apparently in the Aorist Tense). An intermediate pronunciation of /r/ and /z/ is found in Czech.

(2) Chuvash /-l/ vs. Turkic Proper /-sh/;

We have noted several times that the correspondent proto-Bulgaro-Turkic /l/s- liquid seems to survive in modern Khalka Mongolian, cf. the pronunciation of *ula:n* "red" as /ush'a:n, uLa:n/, where /L/ denotes this unique liquid affricate.

Practically speaking, the huge phonological difference between Chuvash and any other Turkic language can be easily observed by comparing almost any Chuvash word, such as 1-10 numbers, to its Turkic Proper equivalent.

### **Bulgaric grammar**

(1) the peculiar plural marker *-sem* in Chuvash (of seemingly unknown origin), absent not only in Turkic but apparently in other Altaic languages. It has been conjectured by a Soviet scholar in a separate article that the Chuvash *-sem*, which rather regularly goes back to *\*-sen*, may only be similar to Kamassian (South Samoyedic) *-saN*. [Kamassian located in the East Sayan Mountains could be in contact with the early Turkic languages, however there is no clear explanation for this phenomenon.]

(2) a peculiar goal-directed case expressed by *-shan, -shen*;

(3) many contracted grammatical forms and a rather simplified grammar in Chuvash (generally typical of contact or "creolized" languages);

### **Bulgaric lexis**

The lexical difference between Chuvash and any other Turkic language amounts to an average of **54.5%** (Swadesh-215, borrowings excluded).

That is roughly equivalent or a little lower than to the lexicostatistical difference between English and any other Germanic language. A similar conclusion has been made by Talat Tekin in [*Türk Dilleri Ailesi (The Turkic Language Family)* // Genel Dilbilim Dergisi, Vol. 2, pp. 7-8, Ankara (1979)], who compared the actual difference between Chuvash and Turkish to the difference between English and German, the latter two,

of course, apart from formally belonging to the same Germanic group and sharing a number of common basic words, are far from being closely related or mutually intelligible.

There is a considerable number of Kazan Tatar lexemes found in the Chuvash basic vocabulary. These lexemes are normally recognizable by their typical non-Bulgarian phonological shape similar to Kazan Tatar or/and the existence of a parallel native word, e.g. *yapâx* "bad", *yeshêl* "green (about grass)", *tinês* "sea", *chechek* "flower", *vârlâx* "seed", *kashkâr* "wolf", *kuyan* "hare", *utrav* "island", *yêbe* "wet" (cf. Tatar *jeben-*, Bashkir *yeben-* "to get wet"), *têrês* "right, correct", etc.

Such common words as *kus'* "eye" and *pus'* "head" may in fact be *too* the Tatar borrowings, taken that they lack the r-ending that is expected in the Proto-Volga-Bulgarian reconstructions *\*xêl* and *\*pul*.

The abbreviated grammar and the considerable number of Kazan Tatar loanwords should be taken into consideration when making conclusions about the origins of Chuvash. Could the early Chuvash be strongly impacted by the Golden Horde language in the past? However, the number of borrowings in Chuvash is hardly much greater than in many other Turkic languages.

### **Bulgarian glottochronology**

Glottochronologically, the separation of a language with the 55% of lexicostatistical differentiation should roughly correspond to anything between 900-1100 BC on the temporal scale. Note that this number has been calculated according to the local temporal calibration, which is neither the standard textbook figure, nor Starostin's method, see again *The Glottochronology of the Turkic languages*.

However, there is some uncertainty concerning this value, because of the logarithmic and statistical nature of the glottochronological principles that makes them prone to errors, *particularly in the cases of standalone languages*. Indeed, the lack of any present-day Chuvash siblings that could allow for a statistical averaging to cancel out any fluctuations, raises doubts about the robustness of this figure. As a result, a relatively small error, which may be due, for instance, to the infiltration of Tatar borrowings, may result in even greater discrepancy when extrapolated beyond the calibration interval, logarithmically modified and projected onto the temporal axis.

At any rate, despite these doubts, the number of about 54-55% is relatively stable, and nearly all the previous estimations performed between 2009-2012 (with the borrowings excluded or included, with different ways to treat synonymy, etc.) have pointed to the early separation of Chuvash, at least as early as 500 BC, but with 1000-1100 BC being a more likely period. Archaeologically, this era of 800-300 BC coincides with the onset

of the early Iron Age in West Siberia, so we may further attempt to support this date by making tentative assumptions about the active use of iron weapons and horse harness during that period, which might somehow have contributed to the Proto-Bulgaric and Proto-Turkic separation.

As it has been mentioned several times, the presence of relatively late dates for the Chuvash separation in other parallel works [Dyachok (2001), Dybo (2006), Mudrak (2009)] is most likely rooted in the application of Starostin's non-logarithmic formulas.

### **Bulgaric history and geography**

In geography, a rather unique European position of Chuvash west of the Urals, a long way from the supposed Turkic homeland near the Altai Mountains (let alone Mongolia, as assumed in certain alternative Urheimat theories) is evident at the very first glance, which again indirectly corroborates the hypothesis of its early separation, given that longer distances presumably correlate with longer migration time.

By the 13th century, Volga Bulgaria must have extended approximately within the 200-km (120-mile) radius from the confluence of the Volga and Kama River. It was probably almost entirely destroyed during the Mongol invasion, making the Volga Bulgarians take refuge in the forested areas of the Volga's right (western) bank, situated within the same 120-mile circle. There, near the forests of Chuvashia, the legacy of Mongolian and Tatar raids must have been less pronounced.

These refugium-type Chuvash settlements in a small area along the Sura (=a tributary of the Volga) are very similar to those of the Mari in the forests and hills of the Volga's left and right bank in the nearby area north of Chuvashia. Unsurprisingly, both ethnicities seem to share certain common ethnological and lexical features (usually seen as Proto-Mari borrowings from Volga Bulgarian).

Consequently, the Chuvash people seem to be those Volga Bulgarians that survived the 13th century's invasion or any later military and cultural interventions by confining themselves to the woodland of Chuvashia and ceding their former territory to the ancestors of Kazan Tatars. The latter ones were clearly first attested in the proximity of the Volga-Kama confluence by Ibn-Fadlan as "al-Bashkird" as early as 922, so their settlement was running almost parallel to that of Volga Bulgarians.

The participation of Kazan Tatar people in the migrational seclusion of Chuvash is obscure. The Kazan Tatars did not necessarily occupy the Volga Bulgarian region by force as part of the Mongolian army in the 1230-40's, rather their settlement in the area of the present-day Tatarstan, though inevitably catalyzed by the disastrous Mongolian invasion, could have resulted from a long and slow migration and linguistic assimilation of Volga Bulgaria extending over a period of many centuries.

It should also be noted that the Chuvash people were first attested in the historical sources only in 1508, and then in 1551, during the rule of Ivan the Terrible and the siege of Kazan by his army. The association of Chuvash with Volga Bulgarians has mostly been the outcome of the historical and linguistic analysis of the 19th century's Turkologists (Kunik, Radlov, Amsharin, etc.) [see the *Brockhaus and Efron Encyclopedic Dictionary* (1906)], however this conjecture is now considered to be well-demonstrated.

**Note:** The ethnonym *Chuvash* is evidently a Tataricized pronunciation of *S'uval*, since the sounds in the former variant may not even exist in Proto-Bulgaric. The city named *Suva:r* is attested near the Etil River (=the Volga), for instance, on the map by Mahmud al-Kashgari (1072-74). He also noted, "*As for the language of Bulgar, Suvar and Bajanak [= Pecheneg], approaching Rum [= that is, from north to south], it is Turkic of a peculiar type with clipped ends.*" [= apparently meaning the rather simplified Bulgaric morphology.]

### **Conclusion:**

The discrepancy between Chuvash and other Turkic languages is so pronounced and its geographical position is so detached from the area of maximum diversification of other Turkic languages that it would be appropriate to separate Chuvash as part of a special Bulgaric taxon within the larger Bulgaro-Turkic supertaxon or family. For most practical purposes, we may assume the date of about 800-1100 BC to be a plausible period for the separation of Proto-Bulgaric from the rest of the Turkic languages.

An important terminological innovation that is suggested in the present study is the usage of the term ***Bulgaro-Turkic*** instead of just *Turkic* for the two major groupings. This terminology modification seems to be reasonable, and arises from the practical need to avoid the continual use of periphrastic expressions like "Turkic Proper", "the Turkic languages outside Chuvash", "the Proto-Turkic homeland excluding Proto-Bulgaric", etc.

### **The Yakutic subgroup**

#### **Where does Sakha actually belong?**

It has been widely accepted since the 19th century's research work, that Sakha, the language of the Yakuts, is almost as distant from other Turkic languages as Chuvash.

Nevertheless, the matter is not that simple. It has also occurred to several researchers that the Yakuts may actually be directly related to other Turkic ethnic groups of Siberia, such as Tuvan, Khakas or Altay.

So instead of positioning Sakha and Dolgan into a stand-alone sub-group, the alternative hypothesis suggests the existence of a "Siberian" taxon which would include most of the Turkic languages east of the Irtysh River line.

Trying to prove the existence of this "Siberian" taxon turns into a complicated Turkological problem. At first glance, Sakha differs drastically not only from any other Turkic language, but also from its closest potential Siberian neighbors. But in other respects, it seems to share with them certain linguistic features that are hard to delineate from common archaisms. Below we will study some of these shared "Siberian" features in detail.

### **Yakutic phonology**

In phonology, the Yakutic subgroup is characterized by the following *local innovations not shared* by any other branches:

(1) the loss of the Proto-Turkic **perhaps aspirated \*sH** as in Old Turkic *sekiz* "eight" > Sakha *aGis*; Old Turkic *sen* > Sakha *en* "you"; Old Turkic *suNok* [N=ng] > Sakha *uNuok* "bone";

(2) the stabilization of **the strongly palatalized Proto-Turkic \*S into an "ordinary" s-**, cf. Chuvash *s'altar* but Sakha *sulus* "star";

(3a) the transition of **the intervocalic -s-, -z- into -h-** as in Old Turkic *qizil* > Sakha *kihil* "red";

(3b) the transition of **-ch- into -X-** as in *biXax* "knife", as opposed to *bichaq* in many other Turkic languages [Baskakov, 1969]. This aspiration is even more pronounced in Dolgan, the northernmost offshoot of Sakha, where the *s-* is converted into the *h-* even in the beginning of the word;

(4) The late development of several **diphthongs**, as in *uon* < *\*on* "ten". "Late" since the vocalism is normally much less historically stable than the consonantism and thus should belong to a relatively recent period;

(5) Various assimilations and dissimilations, which mark the existence of a Proto-Yakutic substrate with strong lenition, which made many original sounds unpronounceable and created the hot-potato effect, such as in the borrowing *pahī:ba* from the Russian */spasiba/* "thanks";

Among notable archaisms, the following features can be listed:

(1) The full retention of the archaic intervocal *-t-* as in *atax* "foot", *xatīN* "birch" probably with some fortition, which is similar only to Tuvan *-d/t-* (where this phoneme is semivoiced), but which is quite unlike the more lenitioned Khakas *-z-*;

(2) The probable retention of the so called "*primary*" long vowels, as in *sa:s* "springtime", *xa:r* "snow", *ti:s* "tooth", which, in other branches, are mostly found in Turkmen and Khalaj, and are often believed to be possible remnants from the Proto-Turkic period.

### **Yakutic grammar**

In grammar, in most respects, Sakha exhibits more grammatical differences than similarities to most other Turkic languages, with the exception of Tuvan, Khakas, Altay, where certain local Siberian similarities have been found.

The following grammatical features in Sakha seems to be unique:

(1) Sakha does not seem to use the negative form similar to *e(r)mes* or *deGil*, which is common in other Turkic languages, but rather the *suox* (after the verbs in the future tense and after the adjectives) and ***buol-batax*** (after nouns) are used instead. The latter seems to be unique among Turkic languages. Cf. *men uchuta:l buol-batax-pīn* "I teacher being-not-am."

**Note:** The Bulgaro-Turkic ***\*bol-*** > Sakha *buol-* is an obvious Nostratic parallel to the English "be", which is present in all of the Bulgaro-Turkic languages.

(2) The loss of the genitive marker;

(3) The usage of ***kini*** "he, she" and *kini-ler* "they" (along with the common Turkic *ol* "that (one)"). The former finds parallels probably only in the Bulgaric *ku* "this, that" and Yugur *ku* "he, she". There exists a hypothesis of its relatedness to Turkish *kendi*, Karakhanid *kendü* "self" (probably going back at least to Ubryatova (1960-80's), a researcher of Dolgan and Sakha (?)), which runs into certain semantic difficulties, though apparently plausible;

(4) The phonologically odd plural pronoun ***ehigi*** (you) with its unique phonological shape, so different both from the conventional *siz* and *seler*;

(5) The unusual comparative case with ***-ta:Gar, -da:Gar, -la:Gar, -na:Gar***. A similar ending for the comparative case is also known in Kimak and Yugur.

On the other hand, the following grammatical features *in nouns and pronouns* seem to be shared with the Altay-Sayan subgroup:

(1) The typical and persistent usage of expressions like *kim-da, kaida-da + a positive verbal construction* denoting indefinite pronouns as in "something does", "somewhere is" and *kim-da, kaida-da + a negative verbal construction* denoting negative pronouns as in "no one did", "nowhere is", etc.

Cf. Sakha *kim-da, hanna-da*; Tuvan *kim-da, kaida-da*; Tofa *qum-ta*; Khakas *kem-de, xayda-da*; Kumandy *kem-de, kaida-da*; Standard Altay *kem-de, \*kaida-da*;

However, this syntactic model is by no means unique to "Siberian", since similar models also exist in Karachay *kim ese da* "someone", *qaida ese da* "sometimes", Tatar *ber-kem (de), (ber) kaida da* and probably elsewhere. In other western Turkic languages, these constructions have mostly been displaced by phrases of Persian origin, therefore this feature is most likely to be a Proto-Turkic archaism, not a Siberian innovation;

(2) The peculiar instrumental case ending in *-nan* shared at least with the Khakas instrumental case ending in *-naN, -neN*. Nevertheless, this feature is evidently a retention, taken that Karakhanid, Old Uyghur, Orkhon Old Turkic and Khorezmian all had a very similar instrumental case with the *(n)in, (n)un, (n)an, (n)ün* marker.

Furthermore, we will provide a brief summary of the Sakha *verbal morphology*:

Notable features of Sakha verbal morphology and their Turkic parallels		
Tense	Sakha	Parallels in other Turkic languages
Imperative 2	<i>bar-iy</i> "please go";	
Imperative 3	<i>bar-ar</i> "go later";	

Tense with <i>-dir-</i>	<i>bar-dar-min</i> "if I go";	Cf. Tofa <i>bar-dir-men</i> "going-am" (Present Continuous)", however with a different meaning (?) Tuvan <i>aytir-a-dir-men</i> "I'm just asking it"; Khakas <i>paz-a-dir-ziN</i> "you're writing"; Altay <i>men bar-a-dir-im</i> "I'm going"; Uzbek <i>yaza-ya-tir-man</i> "I'm writing"; However, Karachay-Balkar and Turkmen dialects are also said to have similar expressions, which makes this grammatical construction a probable archaism.
Optative (apprehensive)	<i>bar-a:ya-min</i> "I think I'd better go (get out)";	Cf. Tofa <i>al-Gay-men</i> "I'd better take it" (Optative), with a little different connotation. A similar marker is also present in Tuvan, Khakas, Altai, Kyrgyz, the languages of the Great Steppe, Cuman-Polovtsian, Karakhanid, Old Uyghur, Khalaj, Yugur, which makes it non-Siberian.
Probability with <i>-tax</i>	<i>bar-daG-iN</i> "you probably go"; <i>as-taG-im</i> "I seem to open";	The ( <i>-dik-</i> ) suffix is present at least in Oghuz-Seljuk and Old Turkic and therefore cannot be Siberian-specific. It seems to be an archaic retention.
Past, Negative with <i>-tax</i>	<i>bar-ba-tax</i> "I have not gone";	Old Turkic ( <i>-maduq</i> ), but not in Siberian Turkic, apparently a retention, as well.
Sporadic necessity with <i>-tax</i>	<i>bar-ar-da:x-pin</i> "Once, I had to go";	Probably, unique to Yakutic.
Future with <i>-iax</i>	<i>bar-iaG-im</i> "I will go", lit. "my going";	May be akin to Tuvan <i>bar-gash</i> "having gone", <i>churu-ash</i> "having drawn". Also, <i>al-gash baar</i> "He will take", <i>kir-gesh kelir</i> "He will come". Apparently, a different usage of the same marker, so it could be Yakutic-Tuvan specific.



Necessity in the future with <i>-iäx</i>	<i>bar-iäh-ta:x-xin</i> "you will have to go";	Probably, unique to Yakutic.
Subjunctive 1 with <i>-iäx</i>	<i>bar-iäx et-iN</i> "if you go";	
Subjunctive 2 with <i>-iäx</i>	<i>bar-iäx e-bi-kkiN</i> "it turns out that you would go"	
Optative-Subjunctive with <i>-iäx</i>	<i>ah-iäx-pit ete</i> "(if) we were opening"; <i>ah-ia suox eti-bit</i> "(if) we weren't opening";	
Usual action with <i>-chi</i>	<i>bar-a:chchi-gin</i> "you normally go";	Probably, akin to <i>-chi</i> in Turkish and other Turkic when denoting professions and occupations, so literally meaning "you are a goer", therefore an archaism with some local additional development.
Positivity	<i>bar-i:hi-gin</i> "you will evidently go";	An archaism, it is also found in Bashkir <i>al-ahi-yim</i>
Probability 2	<i>bar-a:ini-bin</i> "I will probably go";	
Unfinished action with <i>-ilik</i>	<i>bar-a ilik-kiN</i> "you haven't gone yet";	This construction apparently also exists in Khakas ( <i>par-galax-sin</i> ) "you haven't gone yet", Tuvan ( <i>-galak, -qalaq</i> ), Tofa ( <i>-halaq</i> ), Kyrgyz ( <i>-a elek</i> ), possibly Uyghur (?). Also, cf. Tofa <i>alir iik sen</i> "even if I take it". It is the only nearly-certain Siberian isogramme, though, according to Shirokobokova (2005), it seems to be now rarely used in Khakas, Tuvan, absent in Todzin, and archaic in Tofa.
Past unfinished action ("used to")	<i>bar-ar et-im</i> "I used to go";	Present in Oghuz, cf. Turkish <i>var-ir-d-im</i> , therefore cannot be Siberian-specific; a typical retention

Past Tense with <i>-büt-</i>	<i>bar-büt-üm ba:r</i> lit. "my going there is"; <i>bar-bü-ppin</i> "I have gone"; <i>bar-büt etim</i> "I had gone";	A similar suffix ( <i>-miş-</i> ) is present in Old Turkic, Old Uyghur, Khorezmian, Karakhanid, Khalaj, Oghuz-Seljuk, and Tuvan e.g. Tuvan <i>al-bişha:n-men</i> "I'm still getting", but not in other Altay-Sayan languages; an archaic retention. On the other hand, the Great Steppe and Altay-Sayan <i>-Gan</i> past tense is mostly absent in Yakutic.
Past finished action ("once I had to")	<i>bar-büt-ta:x-pin</i> "I had to go once";	
Past, Result	<i>bar-an tur-a-bin</i> lit. "Going, I stand", "I have gone"; <i>bar-an tur-ar-da:x-pin</i> lit. "Going, I stand", "I have gone";	Apparently, similar to the usage of the <i>-Gan-</i> suffix in the languages of the Great Steppe and Altay-Sayan, however the syntactic structure herein is entirely different. Looks like a rather unique Yakutic development.

As it is evident from the table above, most of the shared, allegedly "Siberian", features in verbal morphology are in fact old archaisms found in other branches.

Alternatvely, among the features shared with Orkhon-Oghuz-Karakhanid, and even going back to Proto-Turkic, the following could be mentioned:

(1) The use of *-myt-* / *-byt-* tenses, which are akin to the Old Turkic and Oghuz *-miş-* tenses. These are used only in Oghuz, Salar, Old Turkic, Karakhanid, Khalaj, Cuman-Polovtsian, Uzbek, but not any Altay-Sayan or most Great Steppe languages.

Based on the phonetic similarity of this suffix to Sakha *buol-* that comes from Proto-Turkic *\*bol* "to be" (and the lack of any other specific Yakutic-[Oghuz-Orkhon-Karakhanid] innovations), we can infer that this suffix is most likely an archaism going back to the Proto-Turkic state. Semantically, both the *-büt-* and the *-Gan-* suffixes are in complimentary distribution across the Turkic languages, which basically means that if one is present, the other one is gone or has a different meaning, so apparently, *-Gan-* replaced *-büt-* in Altay-Sayan and most Great-Steppe languages because of the semantic similarity of both tenses.

(2) The use of *-dax-* / *-tax-* / *-daG-* / *-tax-* tenses, which are apparently akin to the Old Turkic and Oghuz-Seljuk *-diG-* / *-tiG-* masdar suffixes.

(3) Cf. the usage of *-er-* instead of *e-*, *i-* as an auxiliary verb "is; to be", cf. Sakha *oGo utuyan erer* "the child is falling asleep" (also similar at least to Khalaj, Old Uyghur and Yugur-Salar), albeit also Sakha *barar etim* "I used to go", where the root of this auxiliary verb *e-tim* is similar to Modern Turkish-Azeri *i-dim* and other Turkic languages.

Most of these features can easily be assumed to be Proto-Turkic archaisms that survived independently in Yakutic and Orkhon-Oghuz-Karakhanid, because presently nothing suggests that they could be a recent innovative development.

On the other hand, there also exist a few unstable Siberian-specific tenses, which can be regarded as suspected Siberian innovations, namely:

(1) The tense with the *-dir-personal ending-* as in *\*bar-dir-men* "maybe I go, if I go", which is actually very typical in the Altay-Sayan languages. However, similar forms have also been found in Turkmen dialects, and are said to be "understandable" by Standard Turkmen speakers, which may be indicative of their existence in Proto-Oghuz.

(2) The tense with the *-a ilik-* construction exists in Altay-Sayan and Kyrgyz (where it is likely to be a borrowing from Altay). However, it seems to have become extinct in most Altay-Sayan languages, so presently it seems to be just a shadow of what it might have originally been, and there are doubts concerning its usage. See [Shirokobokova, N.N. *Otnosheniye jakutskogo jazyka k tyurkskim jazykam Yuzhnoj Sibiri (The relatedness of the Yakut language to the Turkic languages of South Siberia)*, Novosibirsk (2005)]

(3) The use of the *-Gay* participle to show *the optative mood*, as in *bar-a:ya-min* in Sakha and *\*bar-Gay-min* "I'd better go" in Altay-Sayan, whereas in Orkhon-Karakhanid this tense normally expressed the direct future. Nevertheless, such a purely semantic feature is too unstable and could be a naturally occurring independent mutation in meaning both in Proto-Yakutic and Proto-Altay-Sayan;

Most other verbal constructions in Yakutic cannot be found in other Turkic languages, making Sakha verbal morphology rather unique.

### **Borrowings and odd words in the Sakha vocabulary**

Sakha contains lots of words which make one wonder where they could possibly have come from.

In fact, Sakha was described as a mixed tongue at least as earlier as Radlov (1908), who counted that out of 1750 words in a glossary, about 33% were Turkic, 26% were Mongolic, and the rest were of unknown origin.

Presently, we believe that all these borrowings come from at least the four main sources:

(1) Middle Mongolian or the Middle Buryat dialect (pronunciation: /boo-RAHT/).

(2) Evenk (Tungusic);

(3) Russian; as in most "Siberian" languages, the number of Russian loanwords in the abstract and cultural vocabulary is exceedingly high;

(4) an unknown early substrate, most likely of Yeniseian type;

(1) Among potential Mongolic borrowings in the basic vocabulary, one could easily name the following words:

(1) Khakas *sīray*, Altay *chīray*, Tuvan *shīray*, Sakha *sirey* "face" probably from Mongolic, cf. Middle Mongolian *chiray*, Buryat *sharay*. Also, meaning "beauty" in Kyrgyz and Kazakh;

(2) Altay *mechirtke*, Tuvan *merzhergen*, Sakha *mekchirge* "owl" from Mongolic *\*begchergen*, Buryat *begserge* "barred owl";

(3) Sakha *kharba*: "to swim", cf. perhaps Khalkha Mongolian *khayiba*, *khaiva* of the same meaning;

(4) Sakha *moGoy* "snake", cf. Middle Mongolian *moqai*, Khalkha *mogoi*;

(5) Sakha *ergilin* "to turn", cf. Khalkha *ergeG* "turn around";

(10) Sakha *suruy* "to write", *suruk* "letter, mail", cf. Written Mongolian *zhiru-*, Buryat *zura-* "to draw"

The Mongolic origin of some other words is uncertain, though presumable:

(1) Sakha *khallan* "sky", cf. Middle Mongolian *e'ülen* "cloud(s)";

(2) Tuvan *iye*, Sakha *iye* "mother", cf. Khalkha Mongolian *ex* "mother", Evenk *eni:n*;

(3) Sakha *mas* "tree", cf. Khalkha *mod*, Middle Mongolian *mod-un*, Daur *mo:d*, etc., as well as Evenk *mo:*, Nanai *mo:*, Written Manchu *mo:*;

(4) Sakha *bey-em*, Tuvan *bod-um*, Khakas *poz-īm*, Altay *boy-īm* "self", which is probably akin to the Mongolian *bod* and *biye* "body", though this is not necessarily a loanword and could be a retained Altaism;

(2) Some borrowings from Evenk were also found, although in some cases the borrowings could have come the other way around, that is, *into* Evenk, cf.:

Sakha *öydö*: "understand", cf. Evenk *uyde-mi*;

Sakha *oNocho* "boat", cf. Evenk *oNkocho* "wood-board boat", *umurechun* "birch-bark boat";

Sakha *d'i:e* "house", cf. Evenk *d'u*;

Sakha *tīl* "word", cf. Evenk *tīl* "meaning";

Sakha *tarbax* "finger", cf. Evenk dial. *sarbas*;

Sakha *taba* "correct", cf. Evenk *d'abul*;

Sakha *bulta*: "hunting", cf. Evenk *bulta*;

Sakha *seri*: "war", cf. Evenk *kusi:n*, *buleme:chik*, *cherig*, *serI*: (probably, from Sakha into Evenk)

Sakha *örüs* "river", cf. Evenk *birag*, *ene*, *olus* (dialectal), *orus* (dialectal) (apparently, from Sakha into Evenk).

We might conclude that Evenk played some notable role in the formation of Sakha. This is not so surprising considering that Sakha probably acted as a cultural superstratum to Evenk, whereas Evenk, being scattered over the enormous territory of East Siberia, was apparently slowly losing ground to Sakha in the course of the 15th to 20th century.

(3) Russian words are often hard to recognize because they are modified in accordance with the Sakha phonology, cf. the following examples from Swadesh-215: Sakha *chierbe*, Russian *cherv'* "worm"; Sakha *sieme*, Russian *semya* "seed"; Sakha *ba:lki*, Russian *palka* "a stick"; Sakha *bī:l*, Russian *pīl'* "dust"; Sakha *muora*, Russian *mor'e* "sea". This phonological discrepancy implies that other borrowings and archaisms may have also become phonetically unrecognizable. For instance, the following Sakha words of Turkic origin are rather hard to spot at first glance:

Sakha *tīmni* "cold", akin to Karakhanid *tum*, *tumliG* "cold";

Sakha *xaya* "mountain" akin to *kaya* "rock" in most other TL's;

Sakha *ürüN* "white", akin to Orkhon, Old Uyghur, Karakhanid *ürüN*, Khalaj *hirin* "white" (apparently a rare archaism);

Sakha *buruo* "smoke" akin to Old Turkic *bur-* "to boil, evaporate";

(4) The presumable Yeniseian borrowings are particularly interesting.

Sakha *kö* "to fly", cf. Ket *kī* of the same meaning;

Sakha *kötör* "bird", cf. Ket *keNassel*;

Sakha *kini* "he, she, it", cf. Ket *ki*, *kide* [Note that *kini* is normally (probably, according to Ubryatova (1960-80's) explained as being akin to the Karakhanid-Oghuz-Seljuk *kendi* "self", however herein we wonder about a different perspective.];

Sakha *kuttan* "to fear", cf. Ket *koran*, *qoren'*, *qoranai*;

Sakha *söp*, *söptö:x* "right, correct", cf. Ket *sotdas'*;

Sakha *sü:r* "to flow", cf. Ket *sennei*;

It should be noted that Proto-Sakha could not have borrowed directly from Ket, the only living and well-attested representative of the Yeniseian family, but rather from an unknown extinct Yeniseian language. In any case, these presumable cognates are uncertain and are provided herein only as a matter of tentative conjecture.

The presence of an unknown substratum in Sakha probably of Yeniseian origin implies that Proto-Sakha at some point inhabited the Yenisei basin, which is quite reasonable.

There seem to be no noticeable borrowings from Yukaghir among the unidentified words.

### **The few lexical similarities between Sakha and Altay-Sayan**

With only **57%** to Tuvan, **61%** to Khakas, and **56%** to Altay in Swadesh-215 (borrowings excluded), Sakha seems to be a deep-going branch, no doubt of that. It is obviously strikingly different from any other Turkic language. This is because Sakha has many lexical innovations, whose etymology is often hard to explain, and which may in fact turn out to be borrowings from an unknown substrate. However, there seems to exist a number of words common only to "Siberian" languages (= Sakha, Khakas, Tuvan, Altay). Consequently, we should study these suspected examples, attempting to distinguish between archaisms and innovations.

(1) Khakas *izir-*, Tuvan *izir-*, Sakha *itir-* "bite"; however, *isir-* is also found in Turkish, Tatar, Karakhanid and possibly elsewhere, therefore it is an archaism;

(2) Khakas *chiz-*, Tuvan *chod-*, Sakha *sot-* "to wipe"; however, it's akin to Chuvash *sâtâr-*, therefore it is an archaism;

(3) Khakas *köni*, Tuvan *xönü*, Sakha *könö* "straight (as a road)", also cf. Turkmen *göni*. The lexeme is found in many TL's, but this particular meaning only in Siberian Turkic, Altay dialects and Turkmen [see *Sevortyan's dictionary, the V-G-D letters* (1980)]. In any case, apparently, an archaism;

(4) Khakas *xarax*, Tuvan *karak*, Sakha *xarax* "eye". However, *\*qaraq* is also found in Kyrgyz, Old Uyghur and Karakhanid, which makes it a notable but hardly unique Siberian isolexeme. In the meaning "pupil", it is also found in Turkmen and Kyrgyz; the original etymology of this word is evidently "the black part of the eyeball, the pupil". Therefore, apparently, an archaism;

(5) Altay *sogon*, Tofa, Tuvan, Chulym *sogun*, Khakas *sogan*, Sakha *onoGos* "arrow" is usually explained as a cultural borrowing from Samoyedic [Dybo (2007)];

**Note:** *isolexeme* or *isophonolexeme* (introduced herein) is an endemic lexeme, that is a variant of phonological forms and meanings used only within a particular set of languages / dialects in a particular, sometimes rather isolated, territory. For instance, the English lexeme "bad" with its phonological variants /ba:d/, /bæ:d/, etc. and the various typical meanings "not good", "unhealthy", "angry", etc. was originally confined to the dialects of the British Isles and is rather unknown in other Germanic languages. Even if a similar cognate were found in other languages, they would probably have a different meaning or phonological shape. On the contrary, the word "good" is found in many Germanic languages and is hardly a local isolexeme.

On the other hand, the following isolexemes seem to be innovative formations not found outside the supposed "Siberian" subtaxon:

(1) Sakha *sirüt*, Khakas *churt-*, Altay *d'ür-* (*jurtaar*), Tuvan *churtt-* "to live"; obviously, from *\*jurt* "home", "place of pasture", probably innovative, or at least an independent simultaneous semantic formation; note that Sakha included an additional (prothetic?) vowel into the root;

(2) Sakha *siiy-bit*, Khakas *chiziG*, Tuvan *chidig* "rotten" as opposed to *\*chiriq* in most other TL's, including Chuvash; apparently, from *\*J'it-* "to get lost, die, fade";

(3) Sakha *erge*, Khakas *irgi*, Tuvan *ergi* "old" as opposed to *\*eski* in most other TL's;

(4) Sakha *tü:*, Altay *tük*, Tuvan *tük* "wool" instead of the usual *\*Jün*. The original meaning of this word was probably "fluff, fur". Could be coincidental as an independent development;

(5) Sakha *bes*, Altay *mösh*, Tuvan *pösh*, Tofa *bösh* "pine" [Rassadin (1981)];

Another typical "Siberian" feature is preserved in numbers. The "Siberian" 40, 50, 60, 70 are all formed regularly as *\*trt-on*, *\*pesh-on*, *\*alt-on*, *\*s'edi-on*, whereas in any other Bulgaro-Turkic languages, including Chuvash, they retain an irregular structure *\*qrq*, *\*elliG* (evidently from *\*elig* "hand"), *\*alt-msh* / *\*ult-ml*, *\*j'eti-msh* / *\*s'eti-ml*. The regular nouns may have formed in Proto-Sakha due to its stronger isolation from the rest of the Proto-Turkic tribes, and then reborrowed into Altay-Khakas by maintaining trade between Proto-Sakha and Proto-Altay-Khakas, or at least this is the most plausible explanation. In any case, you can see that the number of the purported shared phono-semantic and lexical "Siberian" *innovations* seems to be exceedingly small: we have found only 4-5 words which are difficult to discard outright. It is highly questionable whether this amount could be sufficient to demonstrate the hypothetical Sakha-Altay-Sayan ("Siberian Turkic") common descent. On the other hand, there exist certain words or semantic formations shared not just by Altay-Sayan *but also by the languages of the Great Steppe*, that is, any other languages excluding Orkhon-Oghuz-Karakhanid and Chuvash, e.g.

(1) *\*but* "leg" as opposed to Oghuz-Seljuk *\*but* "thigh"; probably an archaism judging by its presence in other Altaic;

(2) *tün* "night" as opposed to Oghuz-Seljuk *\*dün* "yesterday", but also Chuvash *s'er* "night", *ener* "yesterday"; probably an archaism judging from its presence in Chuvash;

(3) Sakha *aha:*, Khakas *azraan*, Tatar *asharga*, Bashkir *ashau*, Karachay *asharGa* "to eat", whereas in most other TL's the word *ash* is used only to mean "food" (noun); probably a natural semantic development;;

(4) Sakha *xatir-iq*, Khakas *xastir-ix*, Yugur *qazdiq*, Tatar *qayri*, Bashkir *qayir* "(tree) bark", also Tuvan *qazir-iq* "scales, a layer of dirt". Chuvash *xuyâr* "bark" seems to be a borrowing from Tatar. Apparently, an archaism;

These findings could make one wonder whether Yakutic—Altay-Sayan—Great-Steppe may have once constituted a single unity, as opposed to Orkhon-Oghuz-Karakhanid. However, most of these words seem to be archaisms or independent coincidental semantic formations.

### **Unexpected similarities between Sakha and Tofa**



The similarities with Tofa are evident already from the following similar features first discovered by Rassadin in *Morfologiya tofalarskogo yazyka v sravnitel'nom osveschenii (The comparative morphology of the Tofa language)* (1978):

Sakha and Tofa share at least the following features:

- (1) a unique partial case in *-ta/-da*;
- (2) the *-in* ending in the accusative case;
- (3) the adjective ending in *-sīN /gī*, cf. Sakha *-sīN / i*;
- (4) a similar system of onomatopoeic verbs;

However, Tofa is undoubtedly much more similar to the Tuvan subtaxon, than to Yakutic, so no direct genetic unity unifying Sakha and Tofa is supposed to exist. This makes us suspect that *most of the similarities found between Sakha and Altay-Sayan result from a secondary interaction and convergence*. We suspect that *Proto-Sakha may rather have acted as a substrate for Proto-Tofa, so Tofa may have formed when the early Proto-Yakutic speakers switched to Tuvan*.

For the geographical explanation of how this might have happened, see the map below.

### **Conclusions:**

There are drastic lexical differences separating Yakutic from Altay-Sayan (hardly 58% of common words in Swadesh-215), and the majority of Altay-Sayan isolexemes cannot be found in Sakha and vice versa.

Similar considerations refer to the few grammatical and lexical features that Sakha shares with Altay-Sayan *and* the Great-Steppe taxon. The number of these isolexemes and isogramemes is insufficient to make any conclusions concerning their possible unity.

It seems that Sakha just won't fit into the Altay-Sayan subtaxon being pretty much independent. Proto-Sakha was the first to separate from the Proto-Turkic stem at a very early stage, leaving enough time for the Altay-Sayan shared innovations to develop.

Despite the strong Mongolic influence in the vocabulary, Sakha still must retain many archaic features important in the reconstruction of Proto-Turkic.

Moreover, the analysis of borrowings in the basic vocabulary may indicate that Sakha could have initially developed upon an unknown Yeniseian substratum acquired in an unknown area, but most likely when the Sakha were still near the Yenisei basin.

On the other hand, even though the number of possible grammatical and lexical elements shared with Altay-Sayan is rather small and in many cases, there are only tiny traces of innovations, they cannot be discarded outright. It is plausible that Proto-Sakha could have affected the grammar and lexis of Proto-Altay-Sayan leaving a few unexpected common features here and there. That is particularly true of Tofa, that has several shared elements with Sakha, as found by Rassadin (1978-81).

We may conclude that these features shared between Yakutic and Altay-Sayan do not come from their initial genetic relatedness but rather emerge from a secondary contact and convergence. Therefore we may infer that *Proto-Yakutic could have served as a substrate for Proto-Altay-Sayan which later moved along the same route (presumably along the Yenisei) in a secondary migration wave*, thus interacting with Proto-Yakutic and acquiring some of its features.

We may still use the term "**Siberian**" in quotes as a suitable name for the *Sakha plus Altay-Sayan Sprachbund* including any features that they may share either accidentally or due to shared archaisms or as a result of the presumable mutual interaction.

### **How did Sakha actually get there?**

It should be noted that the physical distance from the Altai and West Sayan Mountains to Yakutsk City [or the historical Tuymaada Valley where Yakutsk is located] is just enormous and exceeds 3500 km (2200 miles) in a straight line, being approximately equal to the distance from the Altai Mountains to Chuvashia and Volga Bulgaria along the Volga.

That marks a noticeable curve on the globe and provides an interesting geographical perspective on the matter, making *Sakha and Chuvash look like sort of mirror images of each other*.

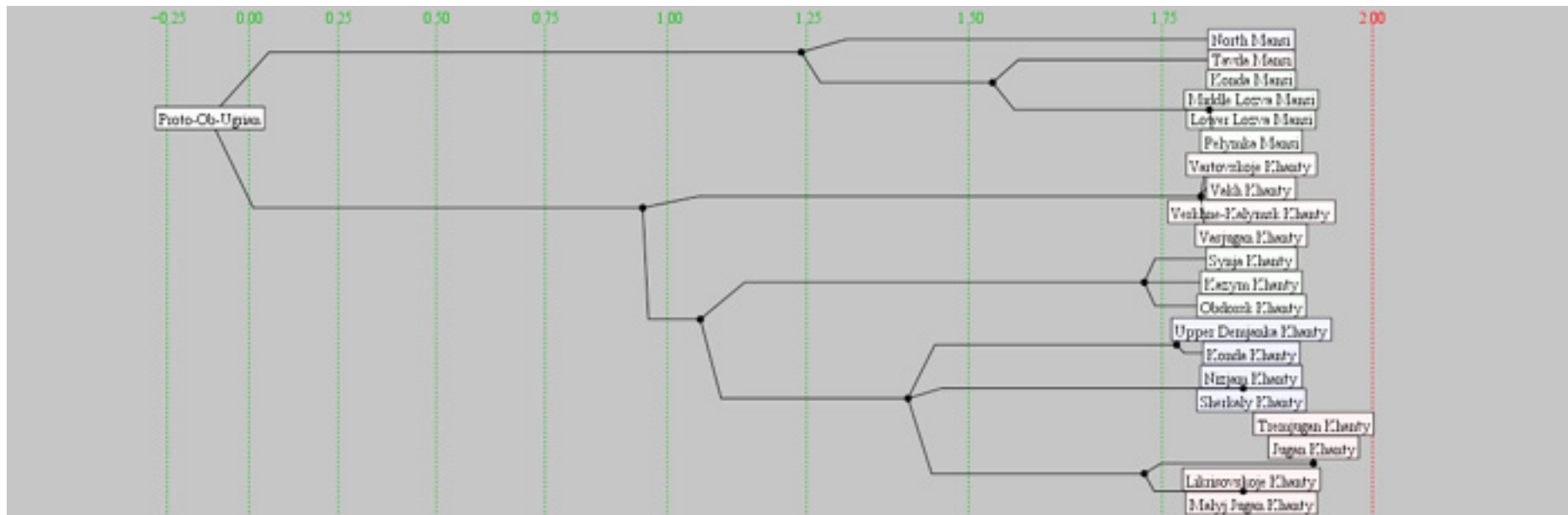
That also poses questions about how and why the Sakha people could have covered that immense distance, when they migrated to the middle Lena. To answer them, we should turn to the consideration of the following points below.

### **The lack of dialectal differentiation within Sakha**

Notably, despite the drastic linguistic differences from other Turkic languages and the gigantic geographic territory it covers, Sakha is rather surprisingly uniform as far as its dialectal differentiation is concerned. It has only one closely related sibling language (Dolgan) and only a few mutually intelligible internal dialects which, for the most part, are reported to differ only in phonology.

This particular point of *absent siblings* makes us infer that the expansion of the Yakuts along the Lena has been a relatively recent event. Otherwise, how can we explain a linguistically uniform expansion over an enormous geographic area extending for three thousand miles? Indeed,

in a similar case with the Khanty language (pronunciation: /HUN-tee, HAHN-tee/) (Finno-Ugric family), in which the Khanty people must have expanded in a similar way over the lower Ob basin in the course of one or two thousand years, we find much stronger linguistic diversification. The dendrogram produced by the group of Georgiy Starostin (2010) confirms the complexities of the Khanty-Mansi internal phylogeny, that consists of multiple language-dialects, so, for all practical purposes, Khanty can presently be viewed as a *taxon*, not a single language. [See [here](#) for details].



The diversification of Khanty-Mansi [Straling database (2010)]

The absence of a similar glottochronological diversification in Sakha as well as the existence of multiple, highly-diversified dialects and lesser-known sub-languages in Khakas, Tuvan, Altai and other "Siberian" Turkic languages of presumably comparable age, the abundance of Mongolian borrowings in Sakha's basic vocabulary, all make us wonder about the peculiarities of Yakutic prehistory.

Naturally, a similar scenario is well-known for Middle English, which has become completely unrecognizable since the Anglo-Saxon times, absorbing many Scandinavian, French and Latin borrowings, but developing very few natural siblings (though its dialectal differentiation is far stronger, and it also has many creole relatives).

It could be surmised that a similar kind of process may have affected Sakha, as well. It seems there could have been a dramatic turning point in Sakha's prehistory that resulted in an ethnological crisis, the inflow of Mongolian loanwords and the extinction of any possible siblings that had existed before that period.

Judging by the lack of dialectal diversification, and the fact that the other in-group sibling languages (besides Dolgan) did not have enough time to develop, that crisis must have occurred during the recent historical past, probably less than a 600-900 years ago.

### **The lack of genetic differentiation in Sakha**

According to Brigitte Pakendorf [Brigitte Pakendorf, *Contact in the Prehistory of the Sakha, Linguistic and Genetic Perspective*, (2007)], "*the genetic results provide clear evidence for the strong founder effect in the Sakha paternal lineage — thus, it is clear that the group of Sakha ancestors who migrated to the north must have been very small*".

The expansion of the Sakha haplotypes (N1c1), found in 90-94% of Yakut population, falls with 95% confidence within the temporal interval between 700 and 1500 CE (idem).

Similar consideration can be found in a different source [Eric Crubezy et al, *Human evolution in Siberia: from frozen bodies to ancient DNA*, BMC Evol Biol. (2010)], which states that the origins of the Yakut male lineages can be traced down to a small group of horse-riders from the Cis-Baikal area (that is, located west of Baikal), which began to spread before the 15th century AD.

This information about the strong bottleneck effect and the existence of just one male progenitor who must have founded all the present-day Sakha clans confirms our hypothesis about the sudden extinction of Sakha siblings in the past.

### **Corroboration from Sakha legends**

According to Sakha legends, the progenitor of all Yakuts was **Elley Bootur**, who was of "Tatar" origin and who fled to the middle course of the Lena, running from "*a great war or persecution*". The word *\*ba: tur* < *\*baGatur* is either a Turkic or Mongolic word for "warrior; strongman; hero" that passed into many languages, hence for instance *Ula:n Ba:tar* "Red Warrior", the capital of Mongolia, or *Yesügei Baatur*, Genghis Khan's father.

Elley Bootur married the daughter of **Omogoy** (or omoGoy, oNohoy, oNoGoy) **Bay**, who had originally lived in the land of Mongols [even though the name's phonology suggests Evenk origin, cf. Evenk *omakta* "new", *emugde* "belly", *oNokto* "nose"], but who had also fled to the north when the wars *during the Genghis Khan rule* (?) broke out. Omogoy Bay had settled down in the delta of the Chara River (a tributary of the Olyokma) near confluence with the Lena about 300 miles from present-day Yakutsk. Alternatively, according to an early version of this legend recorded in the 1740's by Lindenau, Omogoy Bay

lived somewhere along the upper Lena, having fled in that region from Lake Baikal. [*Enciklopedia Yakutii (Encyclopedia of Yakutia)*, Chief Editor: Safronov F. G., Moscow, 2000]

Consequently, our initial hypothesis of mass extinction during the 13th century and a fleeing migration to the north along the Lena continues to find additional support.

The idea that Proto-Sakha tribes could have been persecuted by the Mongols is also partly corroborated by the passages in the *Secret History of Mongols* (1240) [which seems to be the Genghis Khan's personal memoirs written down by a literate scribe in the 3rd person].

The *History* mentions *the genocide of "Tatars" during the early 1200's*. The "Tatars" are said to have been the old enemies of the Mongols, and Genghis Khan's father died three days after paying a visit to a "Tatar" clan feasting in the steppe. These Tatars are said to have lived somewhere near the Onnon and the confluence of the Orkhon and the Selenga, in other words, not too far from the southeastern shores of Lake Baikal, which leads to a conjecture that those "Tatars" could have originally been just an easternmost offshoot of Proto-Sakha.

However, it should also be explained that "Tatar" was apparently just an ancient clan name that could become part of many different ethnicities and could even be used by the Mongols as a misnomer, so we cannot make conclusions about its ethnic or linguistic affiliation just using the name alone. The *History* does not mention which language they spoke or if they could speak a language different from Mongolic.

Yet, in the *Secret History of the Mongols* we also find that Genghis Khan's original name, Temujin, was given because a certain "Tatar" named Temujin-Uge had been captured the day before his birth. This name seems to mean *Temir-ji aGa* "Blacksmith the Elder-Brother", a phrase recognizable in many Turkic languages. Moreover, Genghis Khan's subsequent name may originate from *Tengis Kagan*, where *Tengis* (Turkic "The Sea") is mentioned in the very first lines of the *History*, and presumably refers to Lake Baikal, since there are not too many large lakes in the area. So we may assume that the "Tatars" that lived near the Onon River east of Baikal could indeed have something to do with Turkic tribes.

Even though these inferences are not completely conclusive, they make look the "Tatar"-Kurykan-Sakha connection rather plausible.

### **Positioning Proto-Sakha near Lake Baikal**

Before the time of great crisis, the Proto-Yakuts were probably identifiable with the **Kurykans**, mentioned in one of the Orkhon inscriptions c. 730 as *"üch qurıqan"*, seemingly forming the **Kurumchin archaeological culture** situated near the western shores of Lake Baikal and dated to the 6th-9th century AD. The identification of Proto-Sakha with this culture is a well-known and old hypothesis, based on temporal and geographical considerations and the medieval Chinese records, see [A. P. Okladnikov, *Origins of the Yakut people*(1951)].

The Kurumchin culture, which includes such trades and artifacts as stone walls, sacrificial stones, petroglyphs, agriculture (wheat, rye, millet), iron-making forges, cattle, camel and horse breeding, was focused near the present-day Irkutsk City and around the area of the Murin River (the name itself is

probably akin to Mongolian or Buryat *müren* "river"). The Kurumchin culture could also be found on *Olkhon Island* in Lake Baikal, which is just miles away from the many sources of the Lena basin, including its large upper tributary Kirenga. This proximity of the Lena sources smoothly explains the geographic connection between the northern Yakuts of the middle Lena and their possible southern ancestors, the Kurykans of Lake Baikal.

**Note:** This may also explain why the word *Baikal* seems to be a Turkic hydronym (from *bay* "rich" and *köl* "lake").

### **The distribution of the Buryat and Merkit people**

The present-day distribution of the Buryat people along the western shore of Lake Baikal and the close proximity of modern Buryat to Middle and Khalkha Mongolian suggests that *the Buryat began to arrive in the area of Lake Baikal from Transbaikalia during the early period of the Genghis Khan expansion*. As a result they must have displaced the Kurykan tribes pushing them in the northwest direction.

The *Secret History of the Mongols* tells about the dispersal of the Merkits, a Mongolic clan that, who along with the "Tatars" and the Naimans, were persecuted by the troops of Genghis Khan and his allies in the late 1190's and who tried to escape north by "*entering [the Land] of Bargujin along the Selenga [River]*". In other words they were fleeing towards the eastern shore of Lake Baikal, the area situated between the deltas of the Selenga and the Bargujin, which are the rivers that flow into Lake Baikal at the eastern shore.

As a result the new lands of the Merkits must have been located just 30-50 miles away from the supposed lands of the Kurykans living across Baikal. It is easy to assume that, having been deprived of their cattle and other possessions, and following the domino effect, the desperate Merkits could have attempted an assault at the Kurykans, though these events were naturally outside of the scope of the *History* that mostly tells about Genghis Khan's personal experiences.

Consequently, even though this is entirely hypothetical, we may assume that the Merkits or other neighbouring tribes could have crossed Lake Baikal on ice in winter (only a 20-mile horseback ride) and attacked the Kurykans. They did not even have to apply the extermination policy that Genghis Khan used with the "Tatars", since just destroying winter shelters or taking the cattle away would have led to mass starvation in the Kurykan settlements. Only the few survived by running to the mountains.

This assumption does not explain, however, when and how the Sakha language acquired its Mongolic vocabulary.

The Buryat clan is also briefly mentioned in the *History* as being subject to persecutions, and it is quite plausible that the Buryat, the Merkit and other clans of northern Mongolic tribes have finally contributed to the ethnogenesis of the present-day Buryat people in the vicinity of the southern shores of Lake Baikal and the Trans-Baikalian region, and the presumable exile of the Kurykans.

### **Geography predicts a raft migration from Baikal to Yakutsk**

*How did the Proto-Sakha migrate from Lake Baikal to the present-day area of Yakutsk?*

There seems to be a simple solution to this seemingly complex problem: *the Sakha could have uses a raft or boat migration downstream along the Lena*, so a good portion of this gigantic journey from Baikal to Yakutsk could be accomplished in a relatively short time. This is partly corroborated by one of the legend versions that mentions traveling by raft.

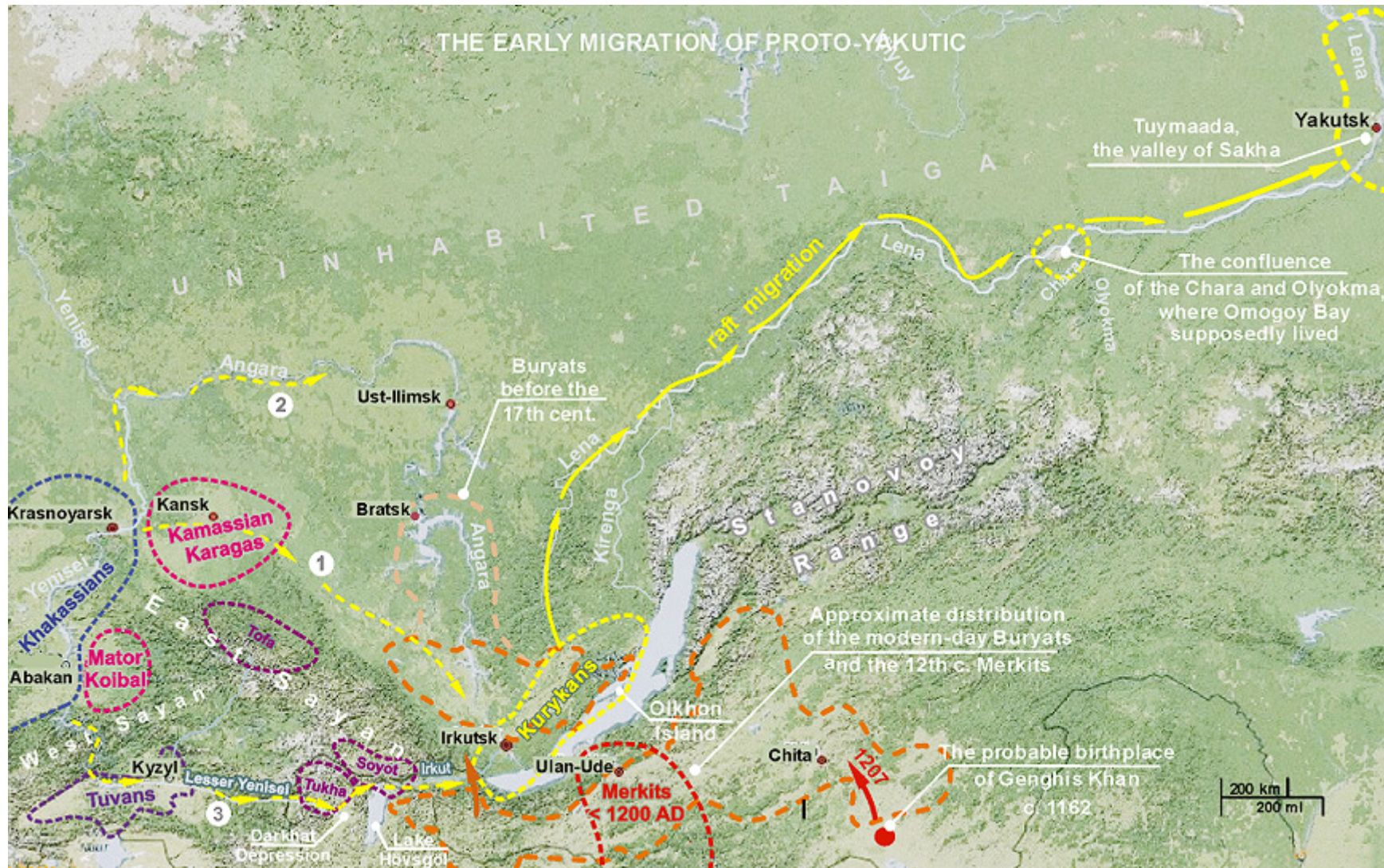
Getting to the Lena River from Baikal is quite easy. The Lena does not have a single source, rather it starts from many small rivers flowing down the western side of the mountain ranges surrounding Lake Baikal, so just a 10-mile walk from the shore across the range will nearly automatically land anyone in the upper Lena River basin — one cannot miss it.

The Tuymaada Valley along the Middle Lena, where Yakutsk City was founded in the 17th century, was known for human settlements since the Bronze Age and even Paleolithic, so evidently the Sakha were not the first to reach this northern territory, and many other ethnic groups could have migrated north using the same route along the Lena.

### **But how did Proto-Sakha even get to Lake Baikal?**

We have established that Sakha demonstrates convergent features shared with the Altay-Sayan and probably some of the Great Steppe languages, all of which are located either along the Yenisei river or further west. So how could Proto-Sakha move from the Yenisei area to the Kurykan settlements at Lake Baikal? And even if they moved to Baikal from an area other than the Yenisei, that migration must still have proceeded from the west, which is getting us back to the same question.

Note that a raft migration *towards* Baikal along the Angara from the west is much less likely, because the Angara flows *from* Lake Baikal, so one has to go upstream in that case.



The early migration of Proto-Yakutic, herein (2011)]

Essentially, there exist three plausible routes from the Yenisei to the Cis-Baikal area [=the area west of Baikal].



### **(1) Across the taiga?**

The Proto-Yakuts may have moved along the East Sayan Mountains and right across the taiga (which includes some of the land belonging to South Samoyedic tribes), that is, roughly along the way of the Trans-Siberian railroad built by the beginning of the 20th century. In a straight line, this potential track would cover a huge distance of over 900 km (550 ml) (from present-day Krasnoyarsk to Irkutsk). It would mostly *cut across* rivers flowing down from the foothills of the East Sayan Ridge, so one would have to know precisely which direction one is taking to get to the destination, given that there is no natural orientation system when traveling *across* a river basin. Therefore such migrations would most likely have had to proceed in a rather random and unsystematic way before the migrants could reach their goal. If this route had actually been taken, we would have presently find many post-Proto-Sakha groups scattered all over the forests between the East Sayan Mountains and the Angara River, which are actually entirely absent.

We should also take into consideration the perils of the taiga travel, such as deep snow in winter, gnat in summer and the evident lack of water as soon as one turns away from the river course. These are obvious reasons why much of this area is still uninhabited up to this day, except for regions with modern roads, railroad tracks and city areas. The attestation of South Samoyedic (Kamassian, Karagas) in the western part of this track, which had supposedly arrived in the area before the Turkic inhabitants and which could probably provide some military opposition to them, equally implies that this territory had most likely been undisturbed until the beginnings of the 17th century. Therefore, we may conclude that the route across the taiga was probably never taken by the Proto-Sakha migrants.

### **(2) Along the Angara?**

Another passable route goes up the Angara River, starting from its confluence with the Yenisei to the Angara's source near the southwestern edge of Lake Baikal. That route is even longer — actually, its length is impossible to calculate precisely because of the many twists and turns of the river's meandering course — but it probably extends for a couple of thousand of kilometers making the potential migrants row hard upstream all the way, with some dense woods and forests along the riverbanks, so neither a natural naval transportation system nor an easily-available shoreline horseback travel could be used for that endeavor.

Winter travel on the ice is more plausible but would probably be hindered by extremely low January temperatures. As in the previous case, no remnants of Turkic tribes were ever found along the Angara or its tributaries. Also note that the many tributaries would tend to divert the migrants away from the initially undetermined destination into even more remote corners of Siberian taiga. We should also keep in mind the possible opposition from the Yeniseian hunting tribes supposedly inhabiting at least some parts of this region. The earliest record of the Russian Cossacks (1620-1630) in the area of Bratsk fortress mention clashes with the "Buryats" and "Tunguses" [=the Evenks] but apparently no Turks / Kyrgyzes / Tatars were spotted in the area, even though the Cossaks had already been familiar with them and should have been able to recognize them.

It is theoretically possible, however, that this type of migration *could have begun* to take place at some point in the past, but probably could not progress very far.

### **(3) The Mongolian track?**

The third possibility is traveling all the way along the upper course of the Yenisei, which would finally land any potential migrants either (1) *in the East Sayan Mountains* — where the Tofa people presently live — (if the potential migrants followed the Greater Yenisei) or (2) *in the Darkhat Depression* with a relatively small lake called Drod-Tsaagan in its center — where the Tsaatan and Soyot people from the Tuvan subgroup presently live and still wander along with their reindeer herds (if the potential migrants followed the Lesser Yenisei).

The Darkhat Depression, the habitat of Tsaatans, is located across the watershed from Lake Hövs-Göl (Khövsgöl), the largest lake of Mongolia, sometimes known as the sister lake of Baikal. Even though, the entire area there is mountainous, traveling along the course of the Lesser Yenisei among relatively sparse Mongolian forests makes it a more viable option. For centuries, this route must have been extensively explored by many reindeer and horse breeding herdsman from Tuva and Mongolia who live in the vicinity, and it is evidently passable.

At the northern edge of Lake Hövsgöl, there is another watershed, beyond which there is the habitat of the Soyots and the source of the Irkut river. As soon as the potential migrants reach the Irkut, it can carry them downstream to the upper Angara in the matter of weeks, and land them all automatically where the present-day Irkutsk City is located, that is, near the area where the Kurykan settlements were attested. The overall track length from Yenisei to Baikal is roughly the same as in the two preceding options — about 1000 km (600 mil), but requiring much less effort, especially in the second half of the journey.

Of course, Tofa curiously shares with Sakha several unique grammatical features, so we have a good confirmation for this hypothesis.

Even more curiously, the self-appellation of the Tsaatans is in fact "Tu'kha" (with an aspirated [t] and a glottal stop in the middle of the word) which is immediately reminiscent of "Sakha". However, this may be a pure coincidence. If it is not, it could be a clan name borrowing or a clan acquisition, when a part of a clan stays to live with another ethnic group.

*Therefore, we may conclude that Proto-Sakha could be a substrate both for Tofa and Tu'kha*, both of which later switched to Tuvan, and this is how the Tofa and Tsaatan (Tu'kha) languages had probably appeared and evolved.

Moreover, the travel through Mongolia could help to explain the Mongolian borrowings in Sakha, though these could also be acquired later from the Proto-Buryats, when the Kurykan people were already near Lake Baikal.

The presence of *the reindeer economy in the Darkhat Depression*, so typical of the Sakha and other North Siberian peoples, is also surprising and may even shed some light on how Sakha and other North-Siberians became reindeer herders. The spread of the reindeer economy from the Sayan Mountains had long been conjectured, but there was no specific mechanism for this process, and the present hypothesis about the movement of Proto-Sakha through the Sayans could shed some light on it, though this complicated matter cannot be discussed here at any length.

In any case, *the Mongolian track seems far more plausible than any other option*, and is well-supported by the lack of geographical obstacles and the presence of ethnographic and linguistic corroborating evidence.

### **Conclusions:**

The analysis of the Sakha dialectal differentiation, genetic makeup and oral history all imply that the Sakha language could have become what it presently is *only after a bottleneck event* that resulted in a dramatic extinction of any sibling clans and their languages.

Before that period, according to the theory created by Okladnikov (1951), as well as judging from the local geography, archaeology and the Chinese and Old Turkic historical records, the Proto-Sakha people may be possibly identified with the *Kurykan* people near Lake Baikal.

The analysis of the *Secret History of the Mongols* (1240) suggests that after the late 1190's the Kurykan Turkic tribes may have possibly been attacked, in the domino effect, by the Mongolic clans, presumably the Merkits and Buryats, which in turn had been pushed from their original settlements by the expanding Mongols of Genghis Khan.

The Kurykans may have tried to escape from the Mongolic invasion by moving north along the Lena River and its southern tributaries in a downstream migration, most likely using simple water transport, such as rafts. This migration down the Lena could have occurred rather swiftly on historical scale.

Before that period, Proto-Sakha had existed in a remote southeastern area, such as the forested ridges adjacent to the western shores of Lake Baikal near the multiple sources of the Lena, possibly even expanding eastwards into Trans-Baikalia and producing some linguistic and genetic offspring east of Baikal. These hypothetical Proto-Sakha groups later became extinct during the Mongol expansion of the early 1200's.

The geomigrational analysis and certain linguistic elements shared with the Altay-Sayan subtaxon, particularly with Tofa (discovered by Rassadin (1981)), suggest that *the Proto-Sakha had migrated into the Lake Baikal area by moving along the upper reaches of the Yenisei River in present-day Tuva*. Proto-Sakha in Tuva must have been displaced there after the arrival of Proto-Tuvan circa 200-300 CE (glottochronological dates) and had to move into the area of the Darkhat Depression and Lake Khövsgöl in northern Mongolia and then migrate down the Irkut River towards Lake Baikal by about 600-800 CE.

### **On the origins of Turkic ethnonymy**

The present article suggests that *nearly all of the Turkic ethnonyms must have had their origins in the names of their clan progenitors*.

The earliest recorded oral Turkic histories, as exemplified by the *Oghuz-Khan Narratives*, written down by Rashid-al-Din (c. 1300), or the *Shajare-i Türk (The Genealogy of Turks)* by Abu al-Ghazi Bahadur (c. 1659), were essentially descriptions of series of legendary events

occurring to Turkic clans and their original male progenitors. Therefore we have a very clear and unmistakable *identification of most Turkic ethnonyms as nothing but patronymic surnames adopted by all the members of that clan.*

For instance, in al-Gazi Bahadur's work, such names as Turk, Oghuz, Uyghur, Kypchak, were clearly and unambiguously associated with male clan founders, including many presumably fictional or real details from their personal lives, which leaves little room for other etymological speculations, e.g.:

*He [Japheth] had eight sons [...] Their names were as follows: **Turk**, Hazar, Saklab, Rus, Ming, Chin, Kemer, Tarykh.*

*But before the Begs gave the answer, the child said, "My name is **Oghuz**."*

*She bore the child in an old (rotten) tree with a hollow. When they told the khan about this, the khan said, "His father died before my very eyes; he has no one to protect him," and so he adopted him. He gave him the name **Kypchaq**. These days a tree with a hollow is called "chypchaq". Humble people, due to slips of tongue, pronounce "kaf" as "chim", thus "Kypchaq" is pronounced as "chypchaq".*

By the same token, Mahmud al-Kashgari (1071-74) says, "*The Turks are in origin twenty tribes. They all trace back to **Turk**, son of Japhet, son of Noah, God's blessing be upon them.*"

Similarly, according to the legend recorded by Ye. S. Filimonov in 1890 [cited in L.V. Dmitriyeva, *Yazyk barabinskikh tatar (materialy i issledovanija) (The language of Baraba Tatars (materials and studies))*, Leningrad (1981)] the progenitor of all the Baraba Tatars was the old man named **Baram** who migrated from a southern land to the north, between the Irtysh and Ob River, where he found plenty of fur animals, birds and fish; there, he had eleven sons — Kelem, Uguy, Uzun, Tulus, Lyubar, Kargal, Kirkach, Choy, Turas, Teren, Baram, — who after Baram's death divided his land into eleven parts (the aymaks). According to Dmitriyeva, these names still mostly correspond to the names of local auls (villages). This legend renders unfounded all the frequent alternative folksy-etymology interpretations of the Baraba name as *barma* "don't go", *baraman* "I'm going", etc. The existence of a specific Baraba clan among other Baraba Tatar clans with different names was confirmed by the demographic data collected and cited by Radlov in 1865 [*Aus Sibirien. Lose Blätter aus meinem Tagebuche (From Siberia: Torn pages from my diary)*, Wilhelm Radloff, Leipzig, 1893].

By the same token, the Khakas legends attribute the origins of the *Khobyy seok* (where "*seok*" means "bone", that is "clan" among the Altay and Khakas people, and which is actually one of the largest clans in the Sagai and Shor ethnicities) to the legendary progenitor named *Kobiy Adas*.

The reason why this evidence has been usually omitted is probably because at some point the scientifically-oriented researchers began to doubt the correctness of mythical factoids described in such legends. However, even if we doubt specific points, there is hardly any reason to doubt the *semantic worldview* in general as adopted by the early Turks and the recorders of these legends.

The early Turkic oral history was documented in a society that reflected the typical male clan social structure, similar to the one described in the Tora and the Quran, where all historical events were likewise often seen as actions of strong and powerful clan forefathers. However, in the course of the 20th century, the original clan structure and the associated ethnographic tradition was almost entirely destroyed and forgotten, consequently a number of folk etymologies and semantically unfounded interpretations concerning the origin of Turkic and Mongolic ethnonyms appeared.

On the other hand, we know full well from historical records that such modern names as *Nogai*, *Uzbek*, *Seljuk* had originally been nothing but personal names, later spreading to the title of a respective dynasty, and then finally to the whole ethnic group or nation.

The expansion from a clan name to an ethnicity or a national name seems to be a common phenomenon occurring with ruling clans that were seen as encompassing the whole large ethnic group.

For instance, it was noted as early as Gerhard Miller (1733-1743):

*"...because the Barabas are, of course, Tatars, as their language shows. Whereas 'Baraba' or 'Barama' is not the name of the whole people, but rather the title of a certain special generation, since other [groups from the Baraba Tatars] also title their generations in a similar way, e.g. Luba, Terenya, Tunus, etc."* [Gerhard Miller, *Istorija Sibirskaja (The History of Siberia)*, Saint-Petersburgh (1750)]

By a "special generation", Miller meant a clan, showing that the Tatars living near Lake Chany originally had many different clans in their social structure, whereas the name Baraba for all of these Tatar clans must have been therefore a recent extension.

By the same fashion, the European surnames also go back to the personal names or aliases of single male individuals, such as *Johnson* to *John*, etc. In both cases, we witness the remnants of *the patriarchal clan structure and the associated patrilineal worldview*.

In the instance of *the Nogai*, we can see that, even though the name originally meant "dog" in Mongolian, there is just as little association with the dogs as in *Bush*, *Green*, *Taylor*, etc. with the respective concepts they represent. Therefore, we may conclude that nearly all the ethnonymic hypotheses or folk etymologies, that attempt to refer a name of a Eurasian ethnic group *directly* to some kind of the real-world phenomena, are usually unfounded, since *nearly all such names originally referred to a personal name or alias of the clan's genetic progenitor or male leader*.

In the Indo-European languages, the original word for "clan" seems to be reflected in the Latin *genus*, Greek *genos*, Irish Gaelic *clann*, Modern English *kin* from Old English *cynn*, Gothic *kunni*, Old Russian *koleno*.

It seems that only after this, we can truly understand the significance of the male haplogroup research conducted in the 1990-2010's. The male DNA markers, just like male surnames, were inherited along the paternal lineage, so they represent the ancient clan markers. And the male clans were pretty much everything to ancient peoples.

In fact, the very usage of the word *adam* for *man* (from Semitic *\*adam*) in most western Turkic languages (e.g. Azeri, Turkish, Tatar, Bashkir, Uzbek, Uighur, Kazakh, Kyrgyz, etc), as well as in Persian, Hindi, Fulani, Indonesian etc., reflects the same tradition of ascribing the descent of the whole ethnic group, even the whole humanity, to one single individual. In this worldview, the history of the whole ethnicity is often seen as an outcome of some action of a legendary ancestor, whose life is poorly understood, with just a few reminiscences surviving in legends, but who presumably passed on his blood to the whole clan, then a confederacy of clans, and finally to the whole ethnic group and even the whole modern nation. (In some cases, however, the name does not go back to the semi-legendary figure himself but rather to that of his father or grandfather, cf. the difference between Seljuk and Togrul Beg.).

Herein, we suggest to name this historiographic conception as *Adamic ethnonymic paradigm*.

It should be stressed that this historiographic worldview is not based on or borrowed from the Abrahamic religions, rather being part of a much older naturally-occurring human tradition.

By the same token, we should infer that the names of other oldest Turkic clans, whose ethnonymic origins have been lost, such as *Kyrgyz*, *Bashkir*, *Kimak*, *Tatar*, *Sakha* and so on, also go back to personal names, rather than any abstract or natural concepts, just because there seems to be hardly any other way of naming clans and ethnic groups in the old Turkic tradition.

For instance, *Kyrgyz* was a surname originally belonging to a male progenitor who received a name or a subsequent alias *Kyrgyz*, probably because of his force, since Turkic verbs *skyr-* "to break" and *kork-* "fear" imply vigor or some fearful action.

Radlov reports (1860's) that the newborn Altayans often received their names from completely accidental events, such as someone entering a yurt with a particular object or something happening shortly before their birth, so we must conclude that trying to find much meaning in clan names will not get us very far. However, leaders like *Temujin*, who got his first name from a Tatar named Temujin-Uge captured the previous day, may have subsequently chosen a more articulate name, e.g. *Tengis Kagan*, from "The Sea" where his ancestors beyond 12 generations had once lived, apparently Lake Baikal.

### **The Altay-Sayan subgroup**

The Sayan-Altay subgroup supposedly includes at least the following languages that belong respectively to the *Tuvan*, *Khakas*, and *Altay* subgroups:

(1) Tuvan, Todzhin, Tofa(lar), Tsaatan, Soyot;

(2) Sagai Khakas (whence Standard Khakas), Kacha Khakas, Kyzyl Khakas, Fuyu Kyrgyz, Mras-Su Shor, Kondoma Shor, Middle Chulym;

(3) Altay-kizhi (whence Standard Altay), Telengit, Teleut, Tuba, Kumandy, Kuu, etc.

Below, we will try to show why this approach to the classification of the local languages seems to be correct.

### **Tofa and Soyot are related to Tuvan**

The fact that Tofa and Soyot are closely related to Tuvan, follows at least from the following evidence.

### **Tuvan, Tofa, Soyot vocabulary**

(1) Dybo's lexicostatistical research (see above);

(2) The fact that most words which are unique to Tuvan (among other TL's) are usually likewise present in Tofa and Soyot, for instance:

Tuvan *chu:(l)*, Tofa *chü*, Soyot *chü* "what?", from Mongolian;

Tuvan *bichi:*, Tofa *biche*, Soyot *biche* "few, little"; "small", also cf. Chuvash *pêchêk*, akin to Mongolian *\*bici-qan* "small";

Tuvan *indi:*, Tofa *indi:* "the other one", apparently, from the Turkic *\*onda* "over there, that one";

Tuvan *uruG*, Tofa *uruG*, Soyot *uriG* "child". of Turkic origin, with the initial meaning "seed";

Tuvan *ashaq*, Tofa *ashiNaq*, Soyot *ashshyaq* "husband", from Turkic;

Tuvan *iye*, Tofa *iGe*, Soyot *i'hê* "mother", probably from Mongolian *ekh*, Buryat *ehe*;

Tuvan *but*, Tofa *but*, Soyot *but* "foot", from Turkic, instead of \**azaq*;

Tuvan *xat*, Tofa *qat* "wind";

Tuvan *xadi:r*, Tofa *qadi:r* "blow (as of wind)";

Tuvan *kesh*, Tofa *ke'sh*, Soyot *ke'sh* "skin", cf. Karakhanid *qas(uq)*;

Tuvan *diNna:r*, Tofa *diNna:r*, Soyot *diNna:(r)* "to hear", from Turkic;

Tuvan *mana:r*, Tofa *mana:r*, Soyot *mana:(r)* "to wait", akin to Khlkha Mongolian *mana-x* "to guard";

Tuvan *eshti:r*, Tofa *e'sht:r* "to swim", also cf. Chuvash *ish-*;

Tuvan *da:ra:r*, Tofa *da:ra:r*, Soyot *da:ra:(r)* "to sew", apparently, a cognate of the normal \**tik* root as in Khakas *tigerge* but with some specific phonological modifications;

Tuvan *xem*, Tofa *xöm* "river";

Tuvan *oruq*, Tofa *oruq*, Soyot *oriq* "road", of Turkic origin, from \**or-* "to dig" [see *SIGTY, Lexis* (2002)];

Tuvan *eqi*, Tofa *e'qqi*, Soyot *eqqi* "good", apparently an archaism, also exists in the Old Turkic *eDgü*, Turkish *iyi*, Karachay-Balkar *igi*, and probably Sakha *üchügey*;

Tuvan *baq*, *baGay*, Tofa *ba'q*, *ba'xay* "bad";

Even though some of these words share parallels with Mongolian, many of them seem to be original Turkic words found mostly only in Tuvan and Tofa, which suggests their close relationship.

### **Tuvan geography**

The geographical relationship between Tuvan and Tofa can be explained in the following way. Initially, the Tuvan people were those Turkic tribes that followed the upper reaches of the Yenisei River into the East Sayan Mountains.



There exist two main sources of the Yenisei, *the Greater Yenisei (Biy-Xöm)* and *the Lesser Yenisei (Ka-Xöm)*. The Tuva's capital Kyzyl is located at their confluence. The many tributaries and sources of the Greater Yenisei lead northeast towards the East Sayan Ridge.

This bordering area between Tuva and Irkutsk Oblast near the West Sayan Ridge is known historically as *Tofalaria*, because Tofa mostly inhabit the East Sayan Mountains, which separate the basins of the Greater Yenisei and the Angara River.

On the hand, the Lesser Yenisei goes east towards Lake Khövsgöl in Mongolia, an area originally inhabited by the *Tsaatans* (in Mongolia) and *Soyots* (in Russia), which, according to Rassadin, the main field researcher of these languages, are closely related to Tofa and Tuvan [see *V.I. Rassadin, O probemakh vozrozhdeniya i sokhraneniya nekotorykh tyurkskikh narodov Yuzhnoy Sibiri (na primere tofalarskogo i soyotskogo)* (2006)]. The Soyots are said to have moved north into Russia from Lake Khövsgöl only 300-400 years ago, though this is mostly based on hearsay evidence from their legends.

Consequently, *Todzin and Tofa must have formed when a part of the Proto-Tuvan tribes moved along the Greater Yenisei (the Biy-Khem), until they reached the forests of the Eastern Sayan Mountains. Whereas, Tsaatan and Soyot must have formed when the Proto-Tuvan tribes moved along the Lesser Yenisei (the Ka-Xöm) towards Lake Khövsgöl in northern Mongolia.*

### **Tuvan hydronymy**

Curiously, the hydronyms of Tyva (Tuva) are clearly and specifically Tuvan, considering they often involve isolexemes or phonetic elements present only in the Tuvan-Tofa subgroup. Cf. *Biche Bash* "small-head (river)", *Ulugan Khöl* "large lake", *Choygan Khöl* "pine lake", *Many Khöl* "Marble Lake", *Chazag* "summer camp (river)", *Kargy* (river) (apparently from *kargaar* "to damn"), *Balyktyg Khem* "fishy river", *Ulug Orug* "big way (river)", *Tashty Khem* "stony river", *Ak Sug* "white water (river)", *Chadan* (apparently from *chada* "step" > river rapid), *Uyuk* "dumbfounding (because of the noise) (a river)", *Chas-Adyr* "springtime fork (spur) (a river)", *Kara Khöl* "black lake", *Khadyn* "birch (lake)", etc. However, the hydronyms quickly change into Mongolian as soon as one crosses Mongolia's and Buryatia's border.

This phenomenon of the local hydronymic continuity is not as common as it may seem and it is probably indicative of the lack of a stable pre-Tuvan substrate in Tuva, and a relatively early occupation of this territory by Proto-Tuvan tribes (about 1500-2000 years ago, which is supported glottochronologically).

## The Khakas languages

### On the origins and usage of the ethnonym Khakas

The term *Khakas* has been introduced only in 1918 during the turmoil of the Russian Revolution, and seems to be nothing but the then-accepted reading of the supposed word "Kyrgyz" in Chinese chronicles, which presumably referred to the Yenisei Kyrgyz people [see the discussion by S. Yakhontov, V. Butanajev, S. Klyashtornyj in the *Etnograficheskije obozrenije* (1992)].

Even today the ethnonym Khakas is rarely used by native speakers, except maybe in formal situations. In fact, Altay and Khakas people have traditionally referred to themselves as just *Tadar(lar)* "Tatars", either because this was the usual name given by Russian Cossacks to nearly all the Turkic peoples in the course of the 17-19th centuries, or because this name could indeed have existed even earlier. The latter point is, however, uncertain.

In any case, the Khakas taxon is subdivided de facto into a number of major dialect-languages, such as *Sagai* (first mentioned in 1311 in Persian records, and then in 1620 in Russian sources), *Kacha* (first attested in 1608), *Kyzyl* (nearly extinct), *Koybal*, *Beltir* (extinct), etc.

The Sagai Khakas people are mostly scattered in rural areas along the foothills of western Khakassia, so pure Sagai is now rarely spoken in cities and seems to be confined to the mountains of the Abakan Range as well as to the area south of the Kuznetsk Alatau Range.

Just like Standard Altay and Standard Crimean Tartar, *the written Khakas is more or less a 20th century's artificial creation based on Sagai*, so most features that are mentioned as typically Khakas in fact refer to the Sagai dialect-language. Since the beginning of the 20th century, when Kacha, located on the planes, has gradually become marginal, Sagai, located in the mountainous areas, can presently be considered as a good sample of a native vernacular Khakas.

Outside the Khakas dialect-languages, the Khakas subgroup includes two other subtaxa — Shor and Chulym — which have long been formally recognized as separate languages, but which too turn out to be small subgroupings:

- (1) **The Shors** (including *Mras-Su Shor* and *Kondom Shor* dialect-languages) mostly live west of the Minusinsk Depression;
- (2) **The Chulym** (including *Middle Chulym* and *Lower Chulym*) live in a completely different region north of the Minusinsk Depression along the Chulym River. Lower Chulym is presently extinct, while Middle Chulym is at the verge of extinction.

### Standard Khakas phonology

Some of the most striking and easily observable phonological features in Standard Khakas, as recorded in a common textbook, in fact come from the Sagai dialect and are not reflected in other Khakas group members (Kacha and Kyzyl). Consequently, these features may result from a recent substratum effect, such as Samoyedic influence.

The following mutations can be regarded as the most typical of the Sagai dialect as compared to other Khakas dialects:

(1) the **-sh** > **-s** mutation as in Sagai Khakas *tas* "stone", *pas* (as in Sakha *ta:s*); but Kachin Khakas *tash*, Shor *tash* "stone", *pash* "head", Tuva, Tofa *t/dash* "stone", *p/ba'sh* "head";

(2) the **-ch** > **-s** mutation as in Sagai Khakas *as-* "open", *sas* "hair", but Kachin Khakas *ach-*, *chach*, Shor *ash-*, *shash*, Tuvan *ash-*, *chash*, Tofa *ash-*, *chesh*; Khakas *aGis* "tree", but Shor *aGash*, Tuvan *iyash*, Tofa *n'esh*;

(3) the **q-** > **x-** mutation in Sagai Khakas as in *xara* "black", but Kachin Khakas *qara*, Tuva *qara*, Tofa *qara*;

It seems that the phonological changes in Standard Khakas and Sagai are relatively recent, whereas *Proto-Khakas* sounded in a much the same way as *Proto-Tuvan* or *Proto-Altay* or many other languages in the region, that is, without these peculiar local phonological mutations.

### **Khakas and Tuvan share few or no exclusive innovations**

Below, we should study the degree of relatedness between Khakas and Tuvan and the plausibility of a separate Khakas-Tuvan proto-state.

### **Khakas and Tuvan phonology**

In phonology, Khakas and Tuvan share the following innovative features:

(1) \***S** > **ch-**, as in Chuvash *s'ichê*, Sakha *sette*, but Tofa *chedi*, Tuvan *chedi*, Khakas *cheti* "seven", and Standard Altay *d'eti* (which is basically pronounced almost the same way as */jeti/*).

Note however, that the \***S-** > **n-** transition is mostly confined to the Khakas subgroup: (1a) *chi-*, *che-* > *ni*, *ne* as Khakas *nimırxa*, Shor *nibirtqa* "egg" as opposed to Tuvan *chuurGa*, but Tofa *n'umurxa*; Khakas *na:x*, Shor *na:q*, but Tuvan *cha:k* "cheek", which sets Tuvan apart from Khakas.

(2) Apparently, a secondary *-w > -G* innovative transition in the final syllable, cf. Tofa *suG*, Tuvan *suG*, Khakas *suG*, Shor *suG*, also Kumandy (a North Altay language-dialect) *su:G / su:*, but Standard Altay *su*: "water". That this is an innovation may be evident from the presumption that *\*suw* must have been the original proto-form.

**Note:** One may be familiar with the Khakas-Tuvan pronunciation of *\*suw* from the name of the *Karasuk archaeological culture*, named after the Karasuk river.

**Note:** The Proto-Turkic *\*suw* and Proto-Bulgarian *\*shuw* (Chuvash *shiv*) "water" is akin to Proto-Mongolic *usun* of the same meaning, evidently from *\*us-sun < \*wus-sun*, where *-sun* is a Mongolic nominative suffix, whereas the root *\*wus-* is most likely Nostratic just like in the English word "water". The same root is also widely distributed in the Uralic languages. Proto-Bulgaro-Turkic seems to go metathetic a number of cases, hence *\*wus > \*suw*.

Therefore, the *w > -G* innovative mutation seems the only phonological feature so far shared by the Khakas-Shor and Tuvan-Tofa subgroupings.

Generally speaking, we have more phonological differences than similarities between Tuvan-Tofa and Khakas-Shor-Chulym. For instance, there are different transitions for the intervocal *-d-*, cf. Khakas, Shor *azaq* "foot", but Tuvan *adaq* "down"; Khakas *xaziN*, Shor *qaziN* "birch", but Tuvan *xadiN*, Tofa *qadiN*.

Moreover, Tuvan-Tofa uses the typical local "Mandarin" system of *weak semi-voiced vs. strong unvoiced plosives* in the consonantism, which is probably derived from the Mongolic languages, and which is also present in many other languages in the region, but not in Khakas.

### **Khakas and Tuvan grammar**

There are very few or basically no innovative features in grammar shared exclusively by the Tuvan and Khakas subgroups, which can be demonstrated in the table below.

<b>The comparison</b>
-----------------------

of Khakas and Tuvan grammatical features		
Grammeme	Tuvan	Khakas
Directive case 1	<i>-che / -zhe</i>	<i>-zar / -zer / -sar / -ser / -nzar / -nzer</i> Rather rare. Also found in Kumandy as <i>-za, -ze-, -sa, -se</i> . This is a differnt ending bearing no relation to the Tuvan equivalent.
Directive case 2	<i>-dive / -duva / -düve / -dīva / -tīve / -tuva / -tüve / -tīva</i>	Shor <i>-taba, -tebe</i> , also Tatar <i>-taba</i> , Kumyk <i>-taba</i> , Kazakh <i>taman</i> , etc, therefore it is not exclusive to the Tuvan-Khakas area.
Differences in the Present Tense	<i>Oyna-p tur</i> "He is playing"; <i>men tur men</i> "I'm standing"; <i>men chor men</i> "I'm walking"; <i>sen chīdir sen</i> "you're lying (on the ground)". The original expression has been preserved in Tuvan and Tofa, whereas the Khakas subgroup developed strong contractions.	Khakas, Shor <i>oyna-p-cha</i> "He is playing" is in fact a standard contraction from <i>*oynap chor</i> . There is some similarity with Tuvan-Tofa, but similar tenses rae present in many other Turkic languages .
The use of a separated pronoun endings as a clitic	<i>men nomcha:n men</i> "I read"	<i>min khiGirgam</i> "I read"; this Khakas construction uses a different ending with a contraction, so they do not match
Differences in the Perfect Tense	<i>men alGan men</i> "I have taken"	Khakas <i>min alGam</i> , Shor <i>men aglGam</i> "I have taken" apparently, with a contraction in the ending.
Differences in the Audative Tense	<i>aytir-a-dir-men</i> "I'm just asking it", "as it turns out I just asked it", the usage of this idiomatic tense is largely similar to the usage of the <i>-mish-</i> tense in Turkish.	Khakas <i>paz-a-dir-zīN</i> "you're writing"; it is identical, however this construction is also shared with Sakha, therefore it cannot be exclusive to the presumable Tuvan-Khakas proto-state.

Differences in the Audative Tense	<i>Kazhan al-chik?</i> "When did he take it, anyway?" <i>Kazhan bar-zhik?</i> "When did he go, anyway?"	Cf. Khakas <i>kil-er-chix-pin</i> "I would come", <i>kil-chiq-ter</i> "Just came". Evidently similar, but it is also attested in Kyrgyz.
Continuous Gerund	<i>kas-pisha:n</i> "(still) digging"; <i>al-bisha:n</i> "(still) taking"; <i>al-bisha:n men</i> "I'm (still) taking"	
Negative Gerund	<i>olur-bain</i> "not sitting, without sitting",	
Unfinished action	<i>al-gizhe-m-che</i> "until (before) I take it"	Khakas, Shor, Altay, Kumyk, Bashkir, Tatar, Uyghur, Karakalpak - <i>gancha-</i> / <i>-genche-</i> , showing unfinished action. But this feature is not exclusive to Khakas-Tuvan.
You (plural)	Tuvan <i>siler</i> , Tofa <i>siler</i>	Khakas <i>sirer</i> , Kumandy <i>sner</i> , <i>snir</i> , Standard Altai <i>slerler</i> , Uyghur <i>silër</i> , Yugur, Salar <i>seler</i> . Not exclusive to Khakas-Tuvan.

So far, we were unable to identify any grammatical features shared exclusively at the level of Khakas-Shor-Chulym and Tuvan-Tofa only. Any similar features are hardly exclusive to these two subtaxa and just seem to point to a different phylogenetic level.

### Khakas and Tuvan vocabulary

With about **72%** for the Tuvan-Khakas pair in Swadesh-215 (as contrasted with the 73% for Turkish-Turkmen and 78% for Azeri-Turkmen), the Tuvan and Khakas languages must be a little further apart than the typical members of the Oghuz subtaxon.

There is hardly any lexicostatistical evidence for Tuvan being any closer to Khakas than to Altay, since we have 72% for Tuvan-Khakas and 69% for Tuvan-Altay.

Most lexical differences between Khakas and Tuvan are due to the large amount of "odd" words in Tuvan and, to a lesser extent, in Tofa. Many of these words turn out to be Mongolic borrowings. Cf. Tuvan, Tofa *chu*: "what" (Khalkha *chu*:); Tuvan *xöy* "many" (Khalkha *xu* "all"); Tuvan, Tofa *urug* "child" (Khalkha *ür*); Tuvan, Tofa *t.ük* "hair" (Khalkha *da:x* "(entangled) hair"); Tuvan *noGa:n* "green", also in Khakas (Khalkha *nogo:n* "green"); Tuvan *mugur* "dull (of a knife)" (Khalkha *molgor*); Tuvan *dayin* "war" (Khalkha *dayin*). However, some of the other Tuvan-Tofa etymologies are much harder to figure out.

### **Khakas and Tuvan geography**

Judging from the geographic perspective, Tuvan is essentially a branch of Proto-Yenisei-Kyrgyz that *migrated further south along the upper reaches of the Yenisei*. Proto-Khakas-Shor-Chulyum originally seemed to inhabit the Minusinsk Depression, whereas Proto-Tuvan-Tofa-Tsataan-Soyot moved further into the Western Sayan mountains, following the course of the Yenisei.

In other words, from the geographic perspective, Khakas-Shor and Tuvan-Tofa (and the closely related language-dialects) are related in the same way as any two ethnicities living in the same river basin. Their mutual contacts, or even the separation from the same stem, should be easily predictable from their geographic position alone. However, one should also take into consideration that both of the subgroups inhabit different mountain valleys. The Khakas subgroup inhabits the Minusinsk Depression, whereas the Tuvan subgroup the Tuvan Depression, both being well-separated from each other by the Western Sayan Ridge.

### **Conclusion:**

After exploring phonological, grammatical and lexicostatistical evidence, we have found no specific innovations shared exclusively by Proto-Tuvan and Proto-Khakas. Furthermore, from the geographic perspective, the two subgroups are separated by the Western Sayan Mountain Ridge. For this reason, the Khakas-Tuvan subgrouping alone — without the inclusion of the Altay subgroup and other related members — seems to be poorly supported.

### **Altay, Khakas and Tuvan form the Altay-Sayan subgroup**

Below, we will study the relatedness of Altay (Turkic) to Tuvan and Khakas trying to demonstrate that, when considered together, these languages form a separate genetically related subtaxon, roughly in the same way as Turkmen, Azeri and Turkish form the Oghuz subgroup.

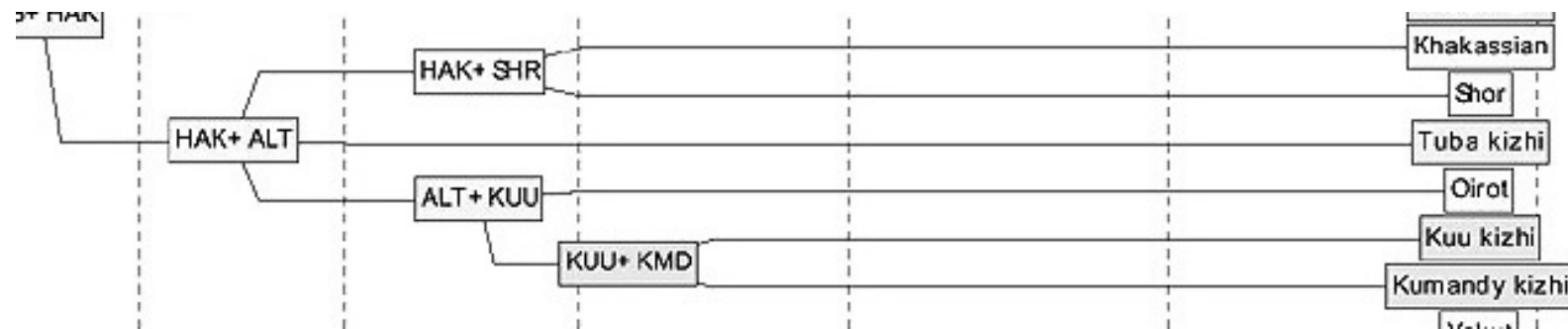
### Altay (Turkic) is not a single language, it is a subtaxon

First of all, as it is well-known today, Altay (Turkic) is not a single language, but rather a complex network of independent languages and dialects. According to Baskakov (1969), the Altay subtaxon should include the following clusters of "dialects":

- (1) Southern: (1a) **Altay-kizhi**, (1b) **Telengit**, (1c) **Teleut**;
  - (2) Northern: (2a) **Tuba**, (2b) **Kumandy**, (2c) **Kuu** (lit. "swan" after the river name) (or Chelkan),
- all of which are probably separate languages.

However, the appellation of the *Altay language* is still widely employed apparently due to traditionalism. This term has been accepted even in Baskakov's works (1952-88), who had done field studies after WWII and written separate books on Kuu (Chalkan) and Kumandy in the 1960-70's.

The strong diversification within Altay (and its relatedness to Khakas) is corroborated by the lexicostatistical study by Anna Dybo (2006).



[Dybo, Anna, *The Chronology of the Turkic Languages and the Linguistic Contacts of the Early Turks* (2006)]

Similar results have been obtained in a phono-morphostatistical study by Oleg Mudrak (2007).



**Note:** the term *Oirot* in the works of Starostin's group members apparently means *Standard Altay* or *Altay-kizhi (Proper)*, which was its official name until 1947.

Moreover, some of the Altay "dialects", such as Kumandy and Kuu (Chelkan), have recently obtained the *de jure* status of separate ethnicities. Curiously, there has even been a sort of small scandal in the press (2011) when two different book authors writing in Kuu argued with each other over which language version should be more correct, so we may surmise there may be some dialectal differentiation even among the speakers of nearby Kuu villages.

The strong diversification within the Altay dialect/languages suggests that Altay (Turkic) peoples have inhabited the Altai Mountains for a long time, presumably at least about a thousand years.

In any case, the Altay Turkic languages are much too peculiar, much too diverse, and were much too poorly studied in the 20th century. Both the Khakas-Shor-Chulym and North-South-Altay subtaxa constitute a rather complex superposition of dialect-languages that could not be explored herein with sufficient elaboration. However, we will attempt to provide a brief argumentation for the Sayan-Altay relatedness below.

### **Altay, Khakas and Tuvan phonology**

It is hard to identify specific phonological features shared exclusively by Altay and Khakas-Tuvan.

Instead, however, we have at least one series of typical contractions shared by Khakas (and partly, Tuvan), Altay, and Kyrgyz. These contractions might have been either archaic or innovative. Cf. the following examples:

(a) as in "liver",

cf. Khakas *pa:r*, Tuvan *pa:r*, Standard Altay *bu:r* / *pu:r*, Kyrgyz *bo:r* "liver", as opposed to Sakha *bīar*, Proto-Kimak-Kypchak *\*bawur*, Chuvash *pôver* < *\*poör* (?) [the Chuvash intervocalic -v- seems to result from the late labialization of narrow vowels], as opposed to Old Turkic *baGir*, probably from Proto-Bulgaro-Turkic *\*Bawir* or *\*Bair*.

(b) as in "bone",

cf. Khakas *sö:k*, Tuvan *sö:k*, Standard Altay *sö:k*, Kyrgyz *sö:k* "bone", as opposed to Sakha *unuoh*, Chuvash *s'ômô*, Old Turkic *süNök* [note that *N* denotes a nasal as the Engl. /ng/], Proto-Kimak-Kypchak *\*süyek*, probably from Proto-Bulgaro-Turkic *\*süNök*.

(c) as in "horn",

cf. Tuvan *mïyïs*, Tofa *mi:s*, Khakas *mü:s*, Standard Altay *mü:s* "horn", as opposed to Chuvash *mây*, Sakha *muos*, Old Turkic *müNüz*, Proto-Kimak-Kypchak and Kazakh-Kyrgyz *\*müyüz*, probably from Proto-Bulgaro-Turkic *\*maNüR* or *\*maiR*.

The details and the direction of these contractions are ambiguous. They seem to be innovative at first, since most contractions are innovative. However, judging by their partial presence in Sakha, and the partial absence from Tuvan, some of them might just as well be quasi-independent mutations or even retentions, so the matter is not entirely clear.

Also note that Kumandy (a North Altay language) exhibits more Khakas features than Standard Altay (Altay-kizhi, "Oiro") [Baskakov (1972)], cf. for instance:

(1) Kumandy *n'*- as in *nimirtka*, cf. Khakas *nimirxa* "egg", but *Jimirtka* (*d'imirtka*) in Standard Altay;

(2) Kumandy *sug* / *su* "water, river" as in Khakas *suG*, Shor *suG*, and Tuvan *suG*, but *suu* in Standard Altay and southern Altay dialects; Kumandy *tag* / *tu* "mountain" as in Khakas *tag*, Shor *taG*, Tuvan *taG*, Tofa *taG*, but *tuu* in Standard Altay and southern Altay dialects;

(3) The Khakas *ch-* instead of the Altay-style *d'-* pronunciation in northern vs. southern Altay dialects, as in *chil* : *d'il* "year"

This affinity has been noted by Baskakov (1969, 1988), who clearly maintained that Northern Altay is rather related to Khakas, whereas Southern Altay to Kyrgyz, which is actually quite illogical, considering the fact that he wrote of Altay as a single language. In any case, it is reasonable to focus on the Southern Altai dialect-languages (Standard Altay, Altay-kizhi, Teleut, Telengit) below, because their relatedness to Khakas seems less obvious.

### **Altay, Khakas and Tuvan grammatical features**

The shared morphological features in Altay-Sayan seem to include at least the following instances:

(1) The use of *choq* after nouns or adjectives (as in "A is not B", or "A is not good") to express negatives instead of or parallel to the standard Turkic *emes*. This feature is typical of many Turkic languages in Siberia. It may also be found in Kyrgyz.

(2) The use of a special contracted form for "you" (plural). Cf. Tuvan *siler*, Tofa *siler*, Khakas *sirer*, Kumandy *sner*, *snir*, Standard Altay *slerler*, Kyrgyz *siler*. Also found in Baraba *assilär*.

(3) The use of a grammeme similar to *bara-dir-min* "I'm going", which also exists in Sakha.

(4) The retention of archaic forms for the past tense 1st person plural (as in "we did"): *-di-bis*, *-di-bis* in Standard Altay and *-di-bis*, *-di-vis* in Kumandy, cf. the innovative *-d'ik*, *-d-uk* in Turkic languages located west of the Irtysh line; this suffix is also reported (rather confusingly) in Standard Altay.

(5) The retention of apparently archaic Optative mood with the *-Gai/-gei-* suffix shared by Sakha, Tuvan, Tofa, Khakas, Standard Altay, Kumandy, Kyrgyz. Even though similar grammemes also exist in other languages, particularly in the Southern supertaxon (see below), they may have a different phonological shape and meaning there (usually the meaning of the future tense).

(6) A special directive case in Kumandy (but not Standard Altay) expressed by *-za*, *-ze*, *-sa*, *-se*, cf. Khakas *-za*, *-zer*, *-sar*, *-ser*, *-nzar*, *-nzer*. Apparently, this feature is quite unique;

### Altay, Khakas and Tuvan vocabulary

Proficient Kyrgyz speakers sometimes report good mutual intelligibility with Standard Altay. Indeed, we have **76%** for Khakas-Altay as opposed to the similar number of **75%** for the Kyrgyz-Standard Altay pairs in Swadesh-215 (borrowings excluded). The distance to any other language from Altay is even greater, with an average of about 70%, or just 69% in the case of Tuvan.

An attempt to find common Altay-Khakas-Tuvan innovative isoglosses produces a bunch of potential lexical innovations:

Basic vocabulary words shared by Altay, Khakas and Tuvan languages				
	Standard Altay	Standard Khakas	Tuvan	

arrow	sogon	soGan	sogun	A cultural borrowing from Ket "soom", probably into Proto-Altay-Sayan (originally, a special kind of a blunt-end arrow used to hunt squirrels, see [Dybo (2006)])
body	neme	nime	et-bot	A possible shared <i>semantic</i> innovation, probably akin to <i>*neme</i> "what".
flea	segertkish	segirtkes	kara-byt	A possible shared innovation
house	tura; <i>also üy</i>	tura; <i>also ib</i>	bazhyN ( <i>also ög</i> (yurt))	<i>Tura</i> is either a shared borrowing from Samoyedic or an innovative noun formed from the verb <i>tur-</i> "stand"
hunger	ach-toro	asta:nï	ashta:nï	But <i>ach</i> , <i>achliq</i> , <i>achtyk</i> in other Turkic. Presumably, a phonological innovation.
young	d'it; d'ash	chi:t; chas	cha:lï < <i>Mong.tsalu</i> :	Cf. the normal <i>*chash</i> in other Turkic, whereas <i>*chiit</i> is akin the to western Turkic <i>*yigit</i> , <i>*Jigit</i> "brave young man", acc. to Starling database. A phonosemantic innovation with the typical Altay-Sayan contraction.
wide	d'albak	chalbax	chalbak, kalbak	A shared innovation in the basic vocabulary; the root also exists in other TLs, but is more common and persistent in this cluster in this particular meaning.
smooth	tüs	tüs	tas	Also, <i>düz</i> in Oghuz-Seljuk, but mostly <i>*tegiz</i> in most languages of the Great Steppe, therefore an archaism.
correct, right	chïn	sïn	shïn	Also, Chuvash <i>chan</i> , therefore probably an archaism, which disappeared in other branches of the TL's.
bad	qomoy	xomay	bagay	Presumably innovative.
root	tazïl	tazïl	t.azïl.	A shared innovation in the basic vocabulary

			<i>Also, Tofa dazil</i>	
bark (n)	chobra	xabix	chövüre:	A shared semantic innovation in the basic vocabulary, probably from <i>*jaburgak</i> (leaf) acc. to Starostin's database
face	chiray	siray	shiray	From Mongolian <i>tsaray</i> from the earlier <i>charay</i> ; however, note that shared borrowings into three languages might not have been borrowed independently from each other.
leaf	pur	pur	p.uru; <i>Tofa pur</i>	As opposed to Kyrgyz <i>Jalbirak</i> , Sakha <i>sebirdeq</i> , etc, which is probably from Proto-Bulgaro-Turkic <i>*SalbirGaq</i> (or a similar proto-form). Either an archaism or innovation.
to laugh	qatqir	xatxir	qatqi	Presumably innovative.
to rub Presumably	jizhar, jizhip siyma:r	chizarGa	t.ürbür	Presumably innovative.
to split (such as wood)	jap	o:darGa	o:ndakta:r	Apparently, absent in other TL's. Presumably innovative.
to scratch (a surface)	jap, cf. tirmaq "fingernail"	tirbax-tir-Ga	t.irbaq; <i>also</i> t.irbaq "fingernail"	Other TL's have the verbal form based on <i>tirnaq</i> "fingernail", but that's phonologically different. Presumably innovative.
to sing	sarinda-, sarna-	sarin sarnirGa	irila:r	A similar word exist in Uygur <i>sayri-maq</i> , Turkmen <i>sayra-mak</i> , but its phonetical shape is different there.
to burn (intr.)	küyer	köyerGe	kivar	Also in Kyrgyz <i>küyü:</i> . Presumably innovative. For the relatedness of Kyrgyz, see below.
to search, look for	bedre:r	ti:irge	t.ile:r	Presumably innovative.

to understand	pilip alar	pilip alarGa	p.ilip alir	Note that the use of the double verbal construction with the <i>-p</i> participle is also very typical of Altay-Sayan and especially Altay languages.
to be over	bozho:r	to:zil pa:rarGa	t.ozar	Presumably innovative, at least in this particular meaning.
sky	teNeri	tigir	t.er:r	This one seems to be semantically archaic, preserving the original meaning of Tengri "sky" instead of "God, Heaven". There exists a hypothesis that this might be a Yeniseian borrowing, akin to Ket <i>tīNgal</i> "tall (about persons)", though that is controversial.
smoke (n)	īsh	īs, tūdün	īsh	With the meaning "soot", it is also known in many other TL's, see [ <i>Sevortyan's Dictionary</i> , Vol. I (1974)], but semantically different in Altay-Sayan.
fat(n)	ūs	ūs	ūs	Unlike the Proto-Bulgaro-Turkic <i>*Su(g)</i> . Presumably innovative.
beast, prey; hunter; to hunt	aN; aNchi; aNdap turar	aN; aNchi; aNnirga	aN; aNchi; aNna:r	Also, noted in Kyrgyz, Kazakh, Karakalpak, Nogai and possibly in Sakha dialects (?) but is much less frequent in these languages. Apparently, from Mongolian <i>aN</i> [ <i>Sevortyan's dict.</i> Vol. I (1974)]. In any case, the verb with the <i>aN</i> -root is Altay-Sayan specific.
to live	d'urt-	churt-	churtt-	Also known in Sakha
tomorrow	jün	taNda	t.a:rta	Presumably innovative
feather	jün	chüg	chüg	Presumably innovative
ice	tosh	puz	t.osh	From <i>ton-</i> "to freeze" > "what is frozen, the frost". Presumably innovative
forest I	arka	agas	arga, arig	Presumably innovative.

forest II	aGash	aGas	arga, arıg	Evidently from <i>*aGach</i> "tree". Also cf. Karachay-Balkar <i>aGach</i> "forest", where it could be an independent development
rain	d'a:sh	naNmır	cha's	Moreover, there are Turkmen <i>yaGish</i> , Azeri <i>yaGish</i> , but these have a different phonological shape and are too far geographically.
mountain	kır	taG	kır	Not found in other TL's in this meaning
war	d'u:	cha:	dayın (Mong.)	Presumably innovative.
island	ortolik	oltirix	ortuluk	A peculiar shared innovative formation from <i>*ortu</i> "middle", cf. Kazan Tatar <i>utrau</i> , Kyrgyz <i>aral</i> , Karakhanid <i>utruG</i> , Turkish and Turkmen <i>ada</i> , etc
(human) back	ucha	ucha	o:rga	Presumably innovative; apparently, not found elsewhere
nose	tumchuq	tumzux	t.umchuq	A possible semantic innovation in the basic vocabulary, probably from a slangy word for "snout", also found in the other TL's, but standard in this meaning only in Altay-Sayan

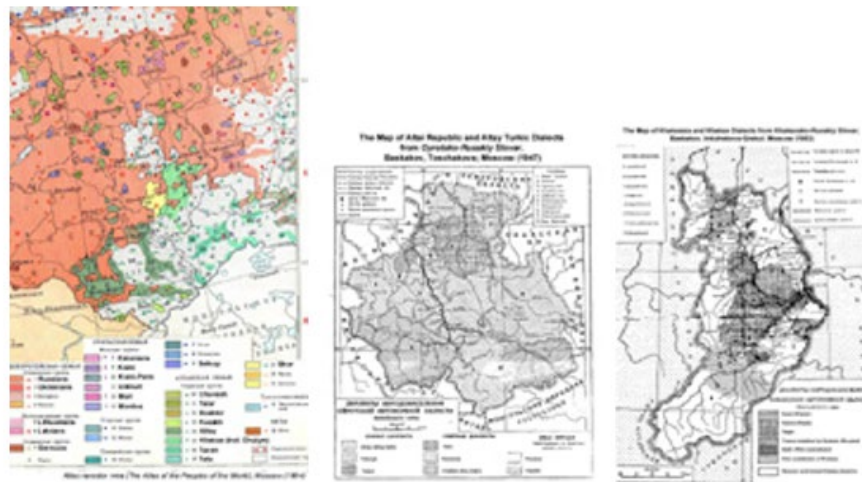
As you can clearly see from the table above, Altay, Khakas and Tuvan share a rather huge number of apparently innovative lexemes, some of which are shared only between one pair of languages, while some of the others are shared across the board. *These isolexemes provide substantial support for the existence of the Altay-Sayan genetic unity.*

As to the reported Altay-Kyrgyz partial mutual intelligibility, it should be noted that most of the lexemes found above are *not* shared with Kyrgyz, setting it apart from the Altay-Sayan languages. Moreover, certain proximity between Altay and Kyrgyz can also be explained by the *considerable linguistic archaism of these two languages and their posterior interaction in the 17-18th century* (see Kyrgyz-Altay isoglosses below).

### **Altay, Khakas and Tuvan history and geography**

The Altai and the Western Sayan Mountains belong to the same mountain system, whereas the Tian Shan is a different matter separated from the Altai Mountains by the basin of the upper Irtysh river. The distance from Lake Issyk-Kul, where Kyrgyz people are presently located, to the Altai Mountains is over 800 km (500 miles). In other words, Altai and Kyrgyz are not geographically connected.

On the other hand, the habitat of the Altai (Turkic) people is very close to the traditional habitat of Khakas, and especially Shor. For instance, the map from the *The Atlas of the World Population* (1964), which supposedly reflects the distribution of ethnic groups during the first half of the 20th century, clearly shows the position of Northern Altai peoples in the direct vicinity of Shor and Khakas.



Old Soviet ethnographic maps of the Altai-Sayan area (1940-60's)

**Note:** The presence of the many unexpected ethnic groups that you can find on the first map, such as Chuvash, Tatar, Mordvins, (Volga) Germans, etc., scattered all over the Altai Krai and Khakassia, is mostly connected with the famine of the 1920's, when there was a mass railroad migration from the Middle Volga to West Siberia, Uzbekistan and other unaffected areas. Presently, most of these ethnic groups must have become ethnically assimilated, at least for the most part, and presumably lost their original languages, though some of them may still exist in the same location.



In any case, we have come to the conclusion that the geographical considerations generally vote for the high probability of Altay-Khakas relatedness and against a readily-available physical connection between Altay and Kyrgyz languages.

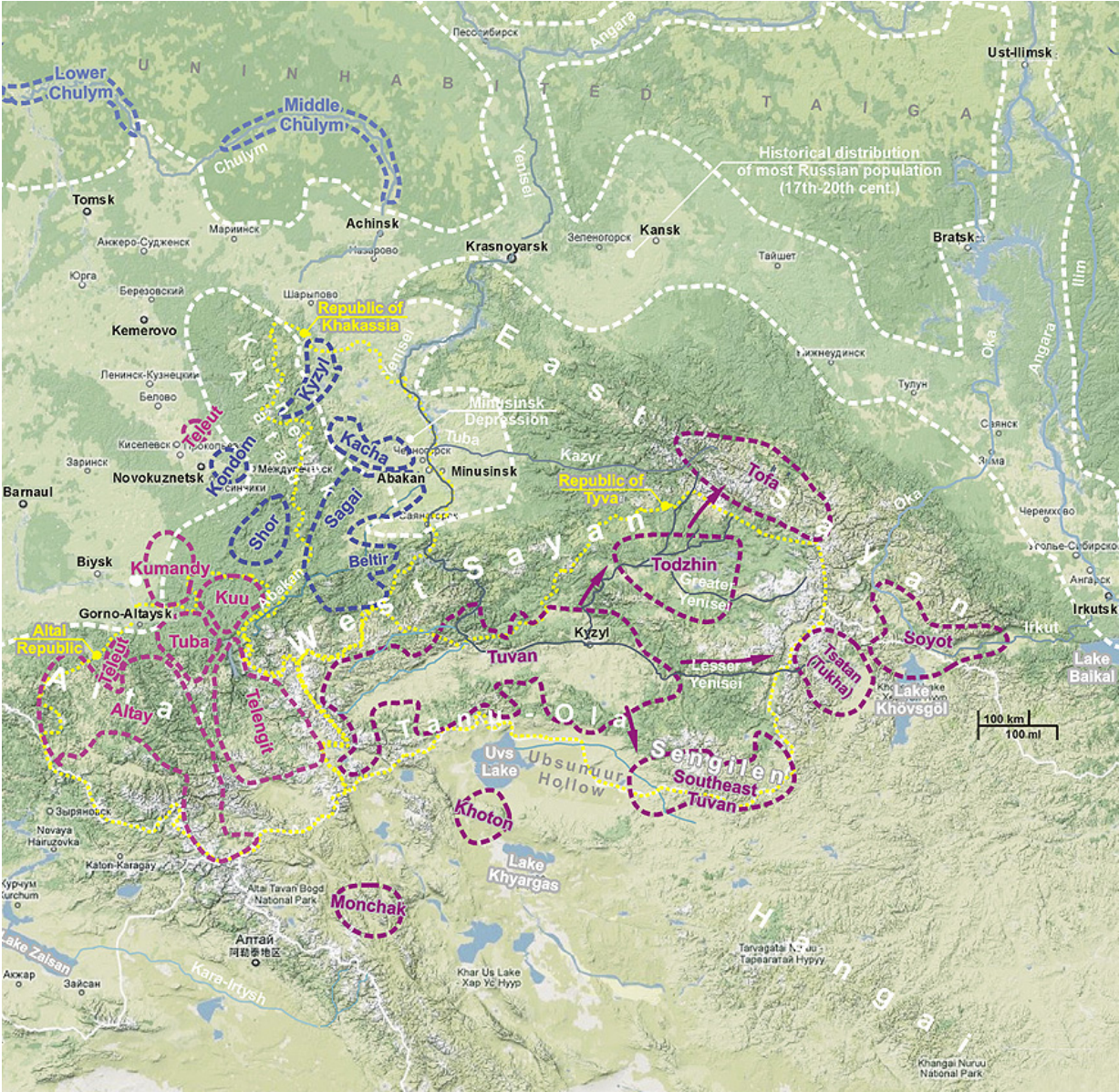
Little is known about the local Altay and Shor history. Curiously, as Radlov mentions about the Shor people in 1861 [*Aus Sibirien. Lose Blätter aus meinem Tagebuche (From Siberia: Torn pages from my diary)*, Wilhelm Radloff, Leipzig, 1893]:

*In vain did I try to exact any historical legends from them [the Mrassu Shors], they could not even name the five ancestors, which any Altayan knows. The 102-year old man could only say that, as he had heard from his father, they had always lived peacefully in this land, and nothing had changed about their way of life except for their faith [=the Orthodox Christianity]; they had always been fishermen, and as far as he could remember, everything stayed the same.*

We may hypothesize that the migration from the Altai to Khakassia or vice versa might actually have proceeded along the Abakan river, which takes source in the Altai Mountains, near the approximate separation area of the Northern and Southern Altay dialects, and which flows through the lands of the Sagai Khakas and Beltir Khakas into the Yenisei River. The Abakan seems to provide an easily available geographic link between the Proto-Khakas and Proto-Altay areas.

**Note:** The interpretation of the Abakan river's name as "bear's blood" is an unlikely option and may represent a folksy etymology, taken that there exists a separate tributary of the Yenisei named *Kan*, as well a number of other rivers in Siberia exhibiting the same root *-kan* presumably meaning "river". Moreover, many other hydronyms in the area do not seem to point towards the Turkic origin, therefore the hydronym *Abakan* may in fact be non-Turkic. More curiously, there exists the *Ubagan* River in the Turgay Vally east of the Urals, but its connection to the Abakan of Khakassia is a mystery.

The ethno-geographical distribution of the Altay Turkic, Khakas and Tuvan subgroups can be summarized in the map below. As in the other similar cases, this distribution mostly reflects the early 20th century situation, when most ethnographic data were collected. By the early 21st century, these areas have shrunk significantly and some dialects (such as Lower Chulym) have even become extinct.



The approximate distribution of the Altay, Khakas and Tuvan peoples by the beginning of the 20th century (2012)

Additionally, the complexity of this geographic distribution leads to a conclusion that the amount of dialectal and linguistic diversification among the members of the Altay, Khakas and Tuvan subtaxa is rather profound and implies at least 1000 years of internal differentiation. By no means do Altay, Khakas and Tuvan presently constitute single, standalone languages.

### **Conclusions:**

Based upon (1) several probable phonological innovations; (2) many shared archaisms in grammar; (3) the large amount of mostly innovative shared isolexemes exclusive to the Altay-Sayan subgrouping, including a well-established lexicostatistical relatedness between Altay, Khakas and Tuvan in Swadesh-215; (4) the geographic proximity and the evident geographic connection between Altay, Khakas and, to a lesser extent, Tuvan languages and dialects;

*we may conclude that the existence of the **Altay-Sayan** proto-state becomes a rather plausible hypothesis.*

Moreover, as lexicostatistical calculations show, there's more proximity between Standard Altay and Standard Khakas on one hand, than between Standard Khakas and Tuvan on the other. We have also shown above that Tuvan and Khakas share no exclusive innovations. These considerations imply that *Proto-Tuvan must have been the first to separate from the Proto-Altay-Sayan stem, whereas Proto-Khakas and Proto-Altay either followed much later or strongly interacted with each other for several centuries, exchanging lexis and phonological features.* At least, the particular relatedness of Kumandy (and reportedly other Northern Altay languages) to Khakas, first noted by Baskakov (1969), can probably be attributed to this later secondary interaction.

During the 2nd millennium CE, a further diversification of Proto-Tuvan, Proto-Khakas and Proto-Altay into smaller languages produced considerable linguistic and dialectal variation in the Altay-Sayan area.

### **The Languages of the Great Steppe**

**Kimak-Kypchak-Tatar, Kyrgyz-Kazakh, and Chagatai-Uzbek-Uyghur seem to form a genetic unity**

According to the present classification, the Turkic languages of the Great Steppe include the following languages and language clusters, among the most typical representatives:

- (1) **Kyrgyz, Kazakh, Karakalpak**, and possibly the extinct dialect of the **Karluks**;
- (2) the spoken medieval **Chagatai**, medieval Sart, modern **Uzbek, Uyghur** and their multiple dialects;
- (3) **Bashkir, Kazan Tatar, Sibir Tatar, Nogai, Kumyk, North Crimean Tatar, Karachay-Balkar**, the unattested **Kimak** dialect, etc.

**Note:** The geographic term *Great Steppe* is used herein to refer to the the western and the largest part of the Eurasian Steppe that stretches from the Altay Mountains to the Black Sea. For more geographical details see also the introduction to *The Proto-Turkic Urheimat & The Early Migrations of Turkic Peoples*.

The Great-Steppe languages seem to share many common elements and are reported to *retain good mutual intelligibility* (subjectively up to 80% in actual speech). Their speakers often get the impression that all of the Turkic languages are very close to each other, even though this impression is in fact connected with the intelligibility of these neighboring languages mostly scattered across the Eurasian steppeland areas and the Tian Shan Mountains in the countries of the former Soviet Union.

In any case, we should suppose that these languages are particularly closely related, and we will try to demonstrate this below.

### **The history and geography of the early Great-Steppe languages**

Apparently, until about 700 AD, all of the proto-members of this presumable supertaxon had occupied the area somewhere near the Irtysh River in the Altay Krai region.

During the rise and fall of the Göktürk-Uyghur Kaganate between the 720-840's, these tribes were affected by the strife with the Göktürks (described in the Orkhon inscriptions), and, probably were compelled to migrate (or allowed to move after the dissipation of Göktürks) from the Irtysh River towards the present-day Kazakhstan, northern Tian Shan, and then deeper into the Great Steppe, though the connection of this migration with the Göktürks-Uyghurs and other details are rather hypothetical and poorly supported.

To establish the earliest known factual migrations, we should first take a look at the earliest attestations of the potential members of this taxon:

(1) The Karluks are reported to migrate from the Altay Mountains to Suyab and establish their confederacy in the Jeti-Su (Zhetisu) by about 760-766 AD. However, virtually nothing is known of this Karluk dialect, and its relatedness to other languages under consideration is purely conjectural. The relatedness of the Karluks to the Kyrgyz is only suggested by their migration to the modern-day Kyrgyzstan and the name's phonology implying superficial similarity with other languages of the Kyrgyz and Kimak origin.

(2) The Tatar clan, presumably forming an important part of the Great-Steppe clans, was first clearly attested, among other Turkic tribes, in the Kul Tegin Orkhon inscription c. 732 in reference to the burial of Bumin Kagan in 552. Judging from the later distribution of the Tatars in the Great Steppe, the Proto-Kimak-Kypchak-Tatar tribes must have been situated along the upper course of the Irtysh River. And indeed, we know they formed their own Kimak Kaganate along the Irtysh after 840 AD.

(3) The Kyrgyz tribes of Kyrgyzstan could have migrated from the Irtysh towards the Jeti-Su region probably after the 840's, that is after the fall of the Uyghur Kaganate (which was essentially the continuation of the Göktürk Empire), when the Yenisei Kyrgyz tribes allegedly sacked the Uyghur capital in Mongolia's Orkhon valley and driven the Uyghurs out of there, establishing their own Kyrgyz Kaganate afterwards. However, the exact details of these events are very confusing, and there are more interpretations in the Russian and Kyrgyz historiography about the origins of the Kyrgyz of Kyrgyzstan than solid facts. An alternative hypothesis suggests that the Kyrgyz had been present in the area between the Tian-Shan and the Altai Mountains since about 200 BCE, when Proto-Turkic tribes and the early "Proto-Central" dialect first appeared in the region [See *The hypothesis of linguistic interaction near Zaisan* below].

Despite the vagueness of the earliest records, *the historical evidence for the Great-Steppe members seems to point to the existence of certain early tribal unities located (1) in the Kulunda Steppe, (2) near the middle-to-upper course of the Irtysh, (3) along the thin strip of land near the upper course of the Irtysh River as it passes through the Altay Mountains flowing from Lake Zaysan.*

*From 200-300 BCE until about 600-800 AD, the early Karluk, Kyrgyz, Tatar and Kimak tribal clans were apparently all situated near this area in the close vicinity of the Kulunda Steppe, Altai Mountains and Lake Zaysan, possibly forming the Proto-Great-Steppe language unity.*

### **The phonology of the Great-Steppe languages**

Most phonological similarities of the three language clusters described above, namely Kimak-Kypchak-Tatar, Kyrgyz-Kazakh and Chagatai-Uzbek-Uyghur, are not exclusive to them, they can also be found in Southern Altay and Oghuz (especially Turkmen), which can probably be attributed to the formation of a local linguistic area.

In other words, besides the Great-Steppe languages being a genetic unity in a strict sense of the word, we may also speak of *the Great-Steppe languages as a Sprachbund* in a broader sense, with some additional ethnicities included in this linguistic area. Some features of this Sprachbund may be present in some of these languages but absent in others. The idea is that most of these Great-Steppe features first arose within the genetic unity, but then spread to other members of the Great-Steppe Sprachbund.

In any case, most languages of the Great Steppe can be characterized by the following phonological characteristics:

(1) A further lenition of **the intervocalic -z- > -y-**: cf. Khakas *azaq*, but Standard Altay and Kumandy *ayak*, Kyrgyz *ayaq*, Kazakh *ayaq*, Chagatai *ayaq*, Kimak-Kypchak-Tatar *\*ayaq*, Oghuz *\*ayaq*. Note that this feature was originally absent from the descendants of Proto-Orkhon-Karakhanid, which preserved a fortified *-d-* or *-ð-*, cf. Orkhon Old Turkic *aDaq*, *adaq*, Karakhanid *aðak* (=the exact pronunciation is uncertain, possibly as a slight interdental /ð/ or an alveolar), Khalaj *hadaq*.

(2) The absence of **the final -G/-g**, as in Standard Altay *tu:*, Kyrgyz *to:*, Kazakh *to:*, Karachay *taw*, Bashkir *taw*, Kazan Tatar *taw* "mountain", but Tuvan *taG/daG*, Khakas *taG*, Kumandy (a Northern Altay language-dialect) *taG*, Oghuz-Seljuk *\*dag*.

(3) Apparently, the **i > e** innovative mutation, as in Standard Altay *eki*, Kumandy *eki*, *ekki*, *iki* (depends on the dialect), Kyrgyz *eki*, Kazakh *eki*, Karachay *eki*, Nogai *eki*, Kumyk *eki* "two", but Tuvan *ihi*, Khakas *iki*, yet Oghuz *\*iki*. Note again that transitions in vowels are often unreliable, lack sufficient historical stability, may emerge independently, or be an areal feature.

(4) A special **voicing pattern** as in Kazan Tatar *sigez* "eight", *tugiz* "nine", Karachay-Balkar *segiz*, *toGuz*, Kyrgyz *segiz*, *toGuz*. Here, the second and third consonants are voiced as opposed to Altay, Kumandy *segis*, *togus*, Khakas *segis*, *toGis*, Yugur *saGis*, *doGis*, Orkhon Old Turkic *sekiz*, *toquz*, Uzbek *sakkiz*, *to'kkiz*.

### **The grammar of the Great-Steppe languages**

(1) The languages of the Great Steppe are characterized by a unique and a very typical shared innovation: the **-d-ik / -d-ik / -d-ük / -d-uk, etc.** Past Tense suffix (1st person, plural) as in "we did" or the **-se-k** in the Subjunctive Mood as in "if we would".

It can be found in some of the Southern Altay language-dialects, Kyrgyz, Kazakh, most Chagatai languages, all of the Kimak-Kypchak-Tatar and Oghuz languages. On the other hand, the suffix is almost entirely absent from the Orkhon-Karakhanid branch [though occasionally present in

late Karakhanid and Khalaj (where it was probably borrowed from Azeri)], "Siberian" Turkic, Yugur, Salar and Chuvash, where the historical archaic *\*-d-imiz* or a synharmonically similar form is used instead in the Simple Past Tense.

**Note:** As a matter of fact, the *\*-d-imiz* suffix is recognizably Nostratic — actually, *-miz* is one of the earliest Nostratic morphemes mentioned by H. Pedersen in his article on Turkish phonology in 1903 — therefore, we may conclude that **-ik / -ïk / -ük / uk, etc** is a later innovation.

(2) At least such languages as Kyrgyz, Kazakh, Chagatai-Uzbek-Uyghur, Karachay-Balkar, Nogai, Karaim exhibit a very odd 3rd person singular *-tï* ending in verbs: cf. Kyrgyz *bara-t*"s/he will go", Kazakh *bara-dï* "s/he is going", Nogai *bara-dï* "s/he goes", Sibir Tatar (Tyumen) *para-tï* "he goes", Uzbek *borap-ti* "s/he is going now", *bara-di* "s/he will go", Uyghur *yazi-du* "s/he, they (will) write".

This pretty striking 3rd person verbal marker, so similar to that of Latin, may make one wonder whether the above-mentioned Turkic languages retained a Nostratic feature. However, it seems to be that this ending is a mere contraction of the common Turkic *-dir, -dir, -dur, -dür, -tïr, -tïr, -tur, -tïr*, used in different connotations in nearly all Turkic grammars and mostly expressing certainty or audative mood. The key to understand how this contraction could have come to life is to realize that the ending *-r* in Turkic Proper is generally unstable and must either transform into a *-z* (according to the law of zetacism) or simply disappear as it happens in modern Turkish dialects, Uyghur and possibly elsewhere. Hence, apparently this *-tïr > -tï > -t* transition in Kyrgyz.

### The vocabulary of the Great-Steppe languages

The lexicostatistical proximity of most Great Steppe languages (except for certain members on the geographic periphery) is quite undeniable and can easily be observed. See for instance, the diagram for the *The Wave Model of the Turkic Languages* above. However, many of these similarities turn out to be archaisms shared with Standard Altay, and sometimes even Khakas, Turkmen and other neighboring languages on the fringe of the Great Steppe, whereas true innovations are harder to detect.

In any case, consider the following lexical and phono-semantical instances, mostly from Swadesh-215, that seem to be innovative because of the absence of these isolexemes in other branches:

(1) Kimak-Kypchak *\*üy*, Kyrgyz *üy*, Kazakh *üy*, Uzbek *öy*, Uyghur *uy*, also St. Altay *öy*, Turkmen *öy* "home" as opposed to Khakas *ib*, Turkish *ev* and a different phonological shape in Tuvan *ög*, Kumandy *ük*. The *\*eb* form is probably more archaic judging from the Korean *chip* and Old Japanese *ipe* "home, house". The *\*öy* word may in fact be more innovative and akin to the Great-Steppe *\*uya*, Seljuk *\*yuwa*, Chuvash *yâwa* "nest", though this latter etymological conjecture does not seem to have been noted anywhere else. [Verified with *Sevortyan's Etymological Dictionary*];

(2) Kimak-Kypchak \**tüye*, Kyrgyz *tö*, Kazakh *tüye*, Uzbek *tuya*, Uyghur *töga*, also Standard Altay *tö*, *tebe*, Turkmen *tüye* as opposed to Khakas *tibe*, Tuvan *teve*, Sakha *taba*, Karakhanid *teve*, Old Uyghur *teve*, Azeri *devä*, Turkish *deve* "camel", Chuvash *teve*. Apparently, this word has undergone innovative phonological modification in Great-Steppe;

(3) Kimak-Kypchak \**may*, Kyrgyz, Kazakh *may*, Uzbek *moy*, Uyghur *may*, also Standard Altay and Altay dialects *may*, Turkmen *may* "fat" (noun), apparently innovative, absent elsewhere. [Verified with *Sevortyan's Etymological Dictionary*];

(4) St. Altay *bet*, Kimak-Kypchak \**bet*, Kyrgyz, Kazakh *bet*, Uzbek *bet*, Uyghur *bet* "face"; apparently innovative. [Verified with *Sevortyan's Etymological Dictionary*];

(5) Kyrgyz *sürt-*, Kazakh *sürt-*, Uzbek *sürt-*, Uyghur *sürt-*, Tatar *sürt-*, Bashkir *hört-*, Karachay-Balkar *sürt-* "to wipe" as opposed to Altay *arla:r*, *archınar*, Khakas *chizrga*, Turkmen *süpür-* "to wipe". Apparently, innovative;

(6) Kyrgyz *oylo:*, Kazakh *oylau*, Uzbek *oyla-*, Uyghur *oyli-*, Tatar *uyla-*, Bashkir *utla-*, Karachay-Balkar –, Turkmen *üyt-*, *pikir et-*, *say-*, as opposed to St. Altay *sanan*, Khakas *saGin-*, "to think, ponder". Apparently, innovative;

(8) Kyrgyz *jirlau*, Kazakh *zhirlau*, Tatar *jirla-*, Bashkir *yirla-*, Karachay-Balkar *jirla-*, as opposed to St. Altay *qozhoNdor*, Khakas *irl-*, Turkmen *sayra-* "to sing". Apparently, innovative;

(9) Kyrgyz *qursaq*, Kazakh *qursaq*, Uyghur *qorsaq*, Tatar *qorsaq*, Bashkir *qorhaq* "belly", as opposed to Oghuz-Seljuk \**qarın*, St. Altay *ich*, Khakas *xarın*, *isti*, cf. Standard Altay *qursak* "pregnant". Apparently, innovative in this meaning. [Verified with *Sevortyan's Etymological Dictionary*];

(10) Kyrgyz *ishku:*, Kazakh *iskilau*, Uzbek *ishqala-*, Tatar *ishqirga*, Bashkir *ishqiu*, Karachay-Balkar *ishirGa* "to rub", as opposed to Oghuz-Seljuk \**sürt(en)*, St. Altay *jizhar*, Khakas *chizarGa*. Apparently, innovative;

(11) Kyrgyz *sürtu*, Kazakh *sürtü:*, Uygur *sürt*, Tatar *sörtörgê*, Bashkir *hörtöü*, Karachay-Balkar *sürterge* "to wipe", as opposed to Turkmen *süpür-* Seljuk \**sil-*, St. Altay *arla:r*, *archanir*. Apparently, innovative;

(12) Kyrgyz *irGitu:*, Kazakh *irGitu*, Tatar *irgitu*, Bashkir *irGitiu* "to throw", as opposed to Uzbek, Uyghur *at-*, Oghuz-Seljuk \**at-*, St. Altay *chachar*, Khakas *tastirGa*, *silerge*. Apparently, innovative;

(13) Kazakh *dala*, Kyrgyz *tala:*, Tatar *dala*, Bashkir *dala*, Uyghur *dala* "steppe, desert". Apparently, innovative but could be a borrowing (?);



(14) Kazakh *dawıs*, Tatar *tawış*, Bashkir *tawış*, Karachay *tawush*, Uzbek *towush*, Uyghur *tawush* "voice". Apparently, is not found elsewhere, therefore probably innovative;

(15) Kazan Tatar *yanGİR*, Bashkir *yamGİR*, Sibir Tatar *yaNGİR*, Nogai *yamGİR*, Karachay *janGur*, Kyrgyz *jamGİR*, Uzbek *yomgir*, Uyghur *yamGur* "rain" is definitely an innovative metathesis from a more archaic *\*jaG-mİR*, which originally seems to have meant "falling water", judging from the fact that the latter word is widely distributed in East Altaic languages as Tungusic *\*mu* "water" and Mongolic *mören* "river", as well as Korean *mul* "water" and even Japanese *mizu* "water". The original variant is attested in all the other Bulgaro-Turkic branches, cf. Chuvash *s'â-mâr*, Sakha *sa-mi:r*, Khakas *naN-mİR*, Altay *jan-mİR*, Turkish *ya:-mur* "rain";

The abundance of archaisms can too contribute to the demonstration, if they come in sufficiently large amounts. Below, there are a few words from Swadesh-215 that seem to be shared archaisms, because of their occasional presence in other Bulgaro-Turkic branches:

(1) Kyrgyz *ötkür*, Kazakh *ötkir*, Uzbek *o'tkir*, Uyghur *ökür*, Tatar *ütken*, Bashkir *ütker*, Turkmen *ötgür* "sharp" as opposed to Karachay-Balkar *jiti*, St. Altay *kurch*, Khakas *chitig* "sharp"; also found in Tuvan, therefore probably a retention;

(2) Kyrgyz *tishte*, Kazakh *tisteu*, Uzbek *tishla-*, Uyghur *chishli-*, Tatar *teshle-*, Bashkir *teshle-*, Standard Altay *tishte*, as opposed to Karachay-Balkar *qab-*, Khakas *izir-* "bite"; a retention;

(3) Kyrgyz *keN*, Kazakh *keN*, Uzbek *keN*, Uyghur *keN*, Tatar *kiN*, Bashkir *kiN*, Karachay-Balkar *keN* "wide", as opposed to Oghuz-Seljuk *genish*, St. Altay *d'albaq*, Khakas *chalbaq*, a retention;

(4) Kyrgyz *qatın*, Kazakh *qatın*, Uzbek *xotun*, Uyghur *xotin*, Tatar *xadın*, Bashkir *qatın*, Karachay-Balkar *qatın* "wife", as opposed to Oghuz-Seljuk *kadın* "woman", St. Altay *üy*, Khakas *ipchizi* "wife", probably a retention;

(5) Kyrgyz *tayaq*, Kazakh *tayaq*, Uzbek *tayoq*, Uyghur *tayaq*, Tatar *tayaq*, Bashkir *tayaq*, Karachay-Balkar *tayaq* "stick", as opposed to Oghuz-Seljuk *chöp*, *chubuk*, St. Altay *agash*, Khakas *agas*, *tayax*, a retention since it is known even in Chuvash *tuya*;

(6) Kyrgyz *soGush*, Kazakh *sogıs*, Tatar *suGış*, Bashkir *huGış* "war", as opposed to Uzbek, Uyghur, Turkmen *\*urush*, St. Altay *d'u:*, Khakas *cha:*, Turkish *savash*. Either archaic or innovative;

(7) Kyrgyz *burulu:*, Kazakh *bu^ru*, Uzbek *bur-*, Uyghur *huri-*, Tatar *borirga*, Bashkir *borolou*, Karachay-Balkar *bururGa*, St. Altay *burilar* "to turn (right, left)", as opposed to Oghuz-Seljuk *\*dön-*, Khakas *aylanarGa*; a retention;

(8) Kimak-Kypchak \**ayt*, Kyrgyz, Kazakh *ayt-*, Uzbek *ayt-*, Uyghur *eyt-*, also St. Altay *ayt-*, Sagay Khakas *ayt-*, Turkmen *ayt-* "to say", though cf. Turkish *ayīt-* "to concern". Apparently an archaism, since it is also found in Sagai Khakas and Sakha as *et* "to tell, to say" and Tuvan *aytīr-* "to explain" and others. However, it is particularly stable as the main verb *fortelling* or *saying* in the languages of the Great Steppe. [Verified with *Sevortyan's Etymological Dictionary*];

### Conclusions:

*A group of tribes inhabiting the Kulunda Steppe and the upper course of the Irtysh River near Lake Zaysan and the Altai Mountains before 600-700 AD* finally led to the formation of the *Kimak-Kypchak-Tatar*, *Kyrgyz-Kazakh*, and *Chagatai-Uzbek-Uyghur* subtaxa. The descendants of these subtaxa are hereinafter referred to as **the languages of Great Steppe, or the Great-Steppe** (super)taxon. Most languages of the Great-Steppe share relatively good mutual intelligibility and many common archaic and innovative isolexemes because of their close linguistic relatedness.

Moreover, some of the languages of the Great Steppe may have additionally affected the development of Turkmen, South Altay, Baraba Tatar and perhaps other geographically related subgroups, in which case we may additionally speak of the **Great Steppe Sprachbund** that includes some languages on the Great Steppe periphery because of the posterior interaction with them.

### Great-Steppe and Altay-Sayan seem to be closer to each other than to Oghuz-Seljuk

We have seen in the discussion above that in some cases the Great-Steppe languages find some similarities with South Altay presumably because of secondary interaction. Below, we will briefly study the features that may genetically relate the Great-Steppe languages to the languages of the Altay-Sayan subgroup at a deeper level.

There are basically two options. If the hypothesis about the Great-Steppe-Altay-Sayan relationship were correct, it would mean that the Orkhon-Oghuz-Karakhanid and Proto-Yakutic branches had been the first to separate from Proto-Turkic Proper, whereas Proto-Great-Steppe-Altay-Sayan split up only several centuries after that. Were it wrong, it would mean that Great-Steppe and Orkhon-Oghuz-Karakhanid should share many common features, whereas Altay-Sayan must have separated early on.

### The grammar of Great-Steppe and Altay-Sayan

(1) The extensive usage of **-Gan-** / **-ken-** in the Perfect Tense instead of the Oghuz-Seljuk **-miş-/-mush-** or Sakha **-bīt-/-mīt-** is rather typical of the Great-Steppe and Altay-Sayan languages. Nevertheless, the **-Gan** suffix is also sporadically present in various direct and indirect functions in Orkhon Old Turkic, Karakhanid, Salar, Yugur, whereas **-miş-** is also known in Cuman-Polovtsian, Uzbek, Tuvan and some other languages. The **-Gan** in Karakhanid and Oghuz-Seljuk is used only in participles and adjectives, not in the Perfect Tense [see for instance *SIGTY. Morphology*. (1988)]. The **-miş-** in Uzbek is evidently inherited from Karakhanid. In Tuvan and Tofa, it has a slightly different meaning of "still doing something", whereas the Perfect Tense is still expressed there with the **-Gan-** / **-ken-** suffix.

Consequently, despite some intermingling, the distinction between the *miş-languages* and *Gan-languages*, which separates Great-Steppe and Altay-Sayan from Yakutic and Orkhon-Oghuz-Karakhanid, altogether seems to be rather sharp and clearly defined.

Since the Oghuz-Seljuk **-miş-/-mush-** or Sakha **-bīt-/-mīt** seem to be an archaism possibly related to the verb *bol-* "to be" and found in the Yakutic branch that must have been the earliest to separate, the usage of **-Gan-** / **-ken-** in the Perfect Tense may turn out to be rather innovative.

Consequently, grammatical considerations seem to point to the Great Steppe and Altay-Sayan relationship.

### **The vocabulary of Great-Steppe and Altay-Sayan**

A few examples of the presumable lexical innovations shared by the Great-Steppe and Altay-Sayan are listed below.

(1) Khakas *omas*, Altay *ötpös*, Tatar *ütmês*, Kazakh *ötpês*, Kyrgyz *ötpögön*, Uzbek *útmás*, Uyghur *ötmes* "dull (of a knife)";

(2) Tuvan *kilir*, Bashkir *kiliu*, Kyrgyz *kilu.*, Uzbek *qilmoq*, Uyghur *qilmak* "to do", whereas in Sejuk-Oghuz this word has been mostly displaced by *etmek* or by *tu* in Chuvash;

(3) Khakas *kiche.*, Altay *keche*, Tatar *kichê*, Bashkir *kisê(ge)*, Kazakh *keshe*, Kyrgyz *keche*, Uzbek *kecha* "yesterday", as opposed to probably more archaic Tuvan *dün*, Uzbektünügün, Karachay *tünene*, Oghuz-Seljuk *\*dün*;

(4) Altay *ölöN*, Tatar *ülên*, Bashkir *ülên* "grass". Moreover, according to Sevortyan's dictionary, cf. Khakas, Kumyk *ölöN* (or similar) meaning "feather grass (=Stipa, one of the most typical kinds of grass in the steppe)"; "*Elytrigia* (type of grass)" in Sakha; "*Carex* (sedge)" in Kyrgyz, Kazakh; "grass" in Uyghur, Uzbek, though modern dictionaries of these languages do not confirm some of the data listed by *Sevortyan's*;

(5) Khakas *köberge*, Altay *köbör*, Karachay *köberge*, Kyrgyz *köbü.*, Uyghur *qaparmak* "to swell (as of a finger, foot)";

(6) Khakas *sörtirge*, Altay *sü:rte:r*, Tatar *söyrêu*, Bashkir *höyrêu*, Kazakh *süyrêu*, Kyrgyz *süyrö:* "to pull (behind oneself)";

(7) Khakas, Tuvan, Tatar, Bashkir, Karachay, Kyrgyz, Kazakh, Altay, \**qol* as opposed to Oghuz-Seljuk \**el*, \**elig*, Sakha *il:i*, Chuvash *alâ*; probably an archaism;

(8) Tuvan *t.ö:*, Khakas *tigi*, Tatar *tege*, Bashkir *tege*, Kyrgyz *tigi* "that (furthest) (adj)", e.g. "that book"; probably a retained archaism, perhaps even of Altaic and Nostratic type;

### **The lexicostatistical considerations for Altay-Sayan and Great-Steppe relationship**

At first glance, lexicostatistically, there is an average distance of about **69%** from Oghuz to Great-Steppe and about **64%** from Great-Steppe to Altay-Sayan (with Tuvan) or 68% (without Tuvan). However, we should take into consideration the mutual lexical exchange among the members of these taxons.

The Great Steppe languages that interacted with the Southern taxon, such as Kimak and particularly Uzbek-Uyghur on one hand, and the Great Steppe languages that interacted with the Altay-Sayan, namely Kyrgyz (see the details in the correspondent chapters). So we are left with Kazakh as the only supposedly "pure" representative of the Great Steppe in our lexicostatistical study. We can also try Bashkir that was confined to the Urals and probably had minimum interaction with Oghuz.

Similarly, we should omit Tuvan from the Altay-Sayan because of the great number of Mongolian borrowings that are hard to detect and that may have infiltrated into the Tuvan list. We should also omit Altay because of its potential interaction with Kazakh, taken that the Altai Mountains form part of the eastern Kazakhstan and there are Kazakh settlements in the Altai.

By the same token, within the Oghuz-Seljuk taxon, we should omit Turkmen because of its potential interaction with Kazakh, Karakalpak and Uzbek, and so we are left only with Azeri-Turkish.

Consequently, the average lexicostatistical distance

(1) for Kazakh and Azeri-Turkish is **66%**;

(1) for Kazakh and Khakas is **68%**;

(1) for Bashkir and Azeri-Turkish is **64%**;

(1) for Bashkir and Khakas is **67%**;

*The resulting difference of 2-3% is very small but the balance now seems to be tipped in the favor of Great-Steppe-Altay-Sayan relationship.*

*In any case, from the lexicostatistical perspective Altay-Sayan, Great-Steppe and Oghuz-Seljuk seem to have separated from each other almost at the same time.*

### **Conclusion:**

It seems that Great-Steppe and Altay-Sayan may be a little more closely related to each other, than either of them is related to Oghuz-Seljuk, Sakha or any other remaining Turkic subgroups. However, the similarities are few and doubts still remain.

We will hereinafter rename this supposed Great-Steppe-Altay-Sayan unity as the **Central** supertaxon for short, because it was geographically located somewhere in the middle between Proto-Sakha and Proto-Orkhon-Oghuz-Karakhanid.

### **The Kyrgyz-Chagatai subtaxon**

As mentioned above, the languages that supposedly belong to this subtaxon are:

(1) Kyrgyz, Kazakh and Karakalpak; (2) medieval Chagatai, modern Uzbek and Uyghur.

### **The history of the Karluks and their bearing on Proto-Kyrgyz-Chagatai**

According to scanty historical records, the Karluks left the Altai mountains circa 665 AD, and migrated towards the Jetti-Su (the Seven Waters region between Lake Balkhash and the Tian Shan Mountains), reaching the Amu-Darya River by about 700. This implies that they may be related to Proto-Kyrgyz-Chagatai originally distributed near the same region (but not at all necessarily).

After the famous Battle of Talas in 751, when the Chinese were defeated by the Arabs and the Arabic supremacy in the region was established, the Karluks were able to form the **Karluk Kaganate** (in 766) by occupying Suyab, the capital of the Western Turkic Kaganate. It was perhaps the political turmoil in the Western Turkic Kaganate, which allowed the Karluks seize power in the Jetti-Su.

The final fall of the Eastern Göktürk Kaganate in 840 left the Karluks in full possession of the Jetti-Su region (the area between the northern Tian Shan and Lake Balkhash). These events must have led to the formation of the Proto-Kyrgyz of Kyrgyzstan (and ultimately, after the 1450's, the Kazakh and Karakalpak languages), though neither the exact details nor the historical relatedness between Karluk and Kyrgyz were clearly documented.

After 840, there could have been a second wave of Kyrgyz migration to the Jetti-Su from the Kulunda Steppe (sources?) that ended political domination of Karluks and finally brought the name of "Kyrgyz" to the present-day Kyrgyzstan (sources?), though the details of this process are still very unclear.

The Chagatai subtaxon, which includes Uzbek, Uyghur and their dialects, is named "Karluk" in Baskakov's classification (see a separate paragraph below). The Baskakov's name "Karluk" for this subtaxon is unacceptable on the same grounds as above: the ethnic affiliation and the exact Turkic dialect spoken by the Karluks are rather obscure. By contrast, the Chagatai origins of Uzbek-Uyghur are well-established.

### **Kazakh is closely related to Kyrgyz**

Before we proceed with the discussion of larger taxa, we will attempt to show the close linguistic relatedness between Kazakh and Kyrgyz, which is an important question for the historiography of Kazakhstan and Kyrgyzstan.

### **The Kyrgyz and Kazakh ethnonymic confusion**

Before the 1920s the Kazakh people were traditionally known as *Kirgizy* "the Kyrgyzes" among Russians. As the often cited anecdote goes [apparently, first mentioned by Kurbangali Khalid (1843-1913)], when asked about their ethnic affiliation, a Kazakh would normally answer something like, "*Men Qazaq-pın*" but corrected by a 19th century's Russian officer, "*What kind of Kazak you are? You're a Kirgiz!*".

The discrepancy is probably due to the frequent application of the ethnonym *Kazak* to the Cossacks of the Polovtsian Steppe and the members of Cossack army. Both are pronounced in Russian as /kazAk/, nearly in the same way as /kazAkh/ "Kazakh", which inevitably resulted in conflation.

As Max Vasmer's *Russisches Etymologisches Woerterbuch* (1950-58) suggests, based on Radlov, who lived among the Kazakh nomads in the 1860's, the original meaning of *Kazak* was "free-lancer, an independent adventurer, soldier of fortune", thus it could have been applied in the medieval period to many different groups of Turkic, Slavic or any other origin. Whether true or not, this interpretation has become generally-accepted.

**Note:** However, this famous Radloff-Vasmer's etymology seems to be rather folksy and hardly corroborated by factual vocabulary. The suffix *-q* seems Turkic indeed. Among roots of similar phonetic shape, there are Turkic *\*qaz-* "to dig", *\*qaziq* "pole", *\*qazan-* "to gain", and

Arabic *qazza:b* "liar", *gazawat* "sacred war", etc. Apparently, there is no reference to a "free-lancer". It is more reasonable to assume that *\*qazaq* had originally been a name of a small clan's leader subsequently lost in history.

The Cossacks of the Ponto-Caspian region must have received their name from the Kazakhs of Kazakhstan via the interaction with the Nogai clans, though there seems to be little specific evidence.

Consequently, to avoid confusion, the Kazakh were officially called *Kazakh Kirgizes*, whereas "the Kyrgyzes of Kyrgystan" — *Kara Kirgizes*. And indeed, in many 19th century's publications, such as Radloff's *Versuch eines Woerterbuches der Tuerk-Dialekte* (1893) printed in German and Russian, Kazakh was formally named Kirgiz (*Kirgizischer Dialekt*), whereas Kirgiz was formally named Kara-Kirgiz (*Kara-Kirgisischer Dialekt*). The Kara-Kirgizskaya Autonomous Oblast was actually the earliest official title of Kyrgyzstan given in 1924.

As to the origins of the ethnonym *Qyrqyz*, there are more wild guesses than well-argued explanations. The name is obviously at least 1500 years old, as it was first mentioned in the Orkhon inscriptions (720's), though probably had existed even earlier. It seems to be the original name applied not only to *Yenisei Kyrgyz tribes*, but also to the members of the *Kyrgyz Kaganate*, and in a broader sense, to most Turkic tribes of the eastern part of the Great Steppe, at least until the Mongol invasion. Moreover, a lake in the Great Lakes Depression in western Mongolia (south of Tuva) was for some reason named *Lake Kyrgyz or Khyargas*, presumably because of the association with the Yenisei Kyrgyz. As a result, *it is actually very difficult to differentiate between the Yenisei Kyrgyz, the Kyrgyz of the Kyrgyz Kaganate, and the early Kyrgyz of Kyrgyzstan*, though all of them seem to be ethnologically different entities.

Phonetically, the word *Qyrqyz* can be associated with *qyr-* "break, smash" or *qorq-* "fear". It seems to be a reduplication, typical of Turkic languages, where the root *\*qyr-qyr* was repeated for emphasis, but the second word-ending *-r* mutated to *-z* according to the law of zetacism in Turkic Proper. The original meaning could therefore be "breaker" (strong warrior).

Most likely, as it has been explained above, the word *Qyrqyz* must have originally been a name or a war alias of a clan progenitor or chief, which later spread to the name of his clan (as in the case with the Seljuks, Noghai, Uzbeks, etc). The event could probably be dated to as early as the beginning of the common era, judging by the action of the zetacism law, thus placing it among the oldest known self-appellations used by the Turkic peoples.

### **Specific phonological features in Kazakh-Karakalpak**

The similarities between Kyrgyz and Kazakh are so many that it is easier to discuss their differences in the first place.

The table below lists some of the phonological differences which seem to have emerged in Kazakh and Karakalpak because of their secondary contact with the Kimak-Kypchak-Tatar languages, particularly Nogai, as well as possibly with some unknown Southern Uralic substratum. By contrast, Kyrgyz seems to be more archaic exhibiting more retentions.

<b>Phonological differences between Kyrgyz and Kazakh-Karakalpak</b>		
<b>mutations and correspondences</b>	<b>Kyrgyz</b>	<b>Kazakh, Karakalpak</b>
<b>ch &gt; sh</b>	<i>chach</i> "hair"	<i>shash</i> , which is similar to Nogai <i>shash</i> and Bashkir <i>säs</i> . The difference can probably be attributed to a local substratum at some point distributed near the Southern Urals.
<b>sh &gt; s</b>	<i>bash</i> "head"; <i>tish</i> "tooth"	<i>bas</i> , <i>tis</i> , which is similar to Nogai <i>bas</i> , <i>tis</i> ; probably due to the action of the same substratum, since similar transitions are also found in Bashkir, and the pronunciation of the Turkmen /s/ is usually interdental, which resembles a comparable mutation.
<b>-0- : -w-</b>	<i>buur</i> "liver"	<i>bawir</i> ; similar to Kazan Tatar <i>bawir</i> , Bashkir <i>bawir</i> , Nogai <i>bawir</i> , Karachay <i>bawur</i> . Apparently from the interaction with Nogai.
<b>-0- : -y-</b>	<i>söök</i> "bone"	<i>süyek</i> ; similar to Kazan Tatar <i>söyek</i> , Bashkir <i>höyäk</i> , Nogai <i>süyek</i> , Kumyk <i>süyek</i> , Karachay <i>süyek</i> ; the -y-formation in this word is not found elsewhere and seems to be an innovative feature that must have come from the Kimak languages, apparently Nogai
<b>-u- : -ï- in suffixes</b>	<i>kuyruk</i> "tail"	<i>qyriq</i> ; similar to Kazan Tatar <i>qoyriq</i> and Nogai <i>qyriq</i> . This is an innovative feature that must have come from Nogai, considering that most TL's have -u- in the 2nd position, see the Starling database.

Also cf. a similar table for Kimak languages (below).



Consequently, we can see that the phonological differences between Kyrgyz and Kazakh-Karakalpak are also shared by some of the Kimak languages that were part of the Golden Horde, particularly the nearby located Nogai. Such phonetic evidence probably led Baskakov to believe that Kyrgyz and Kazakh are not even closely related, and Kazakh should be regrouped with Nogai.

However, judging from the good lexical matches between Kyrgyz and Kazakh that were not measured by Baskakov, this is clearly not the case. Rather, the purported relatedness between the Kimak languages and Kazakh must result from the many shared archaisms and *a few secondary changes in Proto-Kazakh-Karakalpak which came from a posterior interaction of the early Kazakh with the languages of the Golden Horde, specifically and most likely the early Nogai.*

### The grammar of Kyrgyz and Kazakh

Both *Kyrgyz and Kazakh* a great number of archaic features, many of which are also known to exist in the Altay-Sayan Turkic languages. As far as the innovative elements are concerned, Kyrgyz and Kazakh seem to exhibit the following grammatical elements:

(1) Both Kyrgyz and Kazakh use the typical 2nd person plural pronoun, apparently absent from other branches, cf. Kyrgyz *sizder, siler*; Kazakh *sizder, sender*.

(2) A rather unique type of the instrumental case, cf. the Kyrgyz *menen* e.g. *qol menen* "with the hand", Kazakh *-men, -pen, -ben*; also *menen*. Although this feature is probably archaic, taken that *\*menen* is also known in certain dialectal variations of standard languages, such as Eastern Bashkir or Sagai Khakas.

An even greater number of grammatical traits is simultaneously shared with Chagatai-Uzbek-Uyghur languages (see below).

However, beside the similarity, there is also some notable discrepancy in grammatical usage and morphology:

	Kyrgyz	Kazakh
pronouns in the ablative, e.g. "from me"	men-den	men-en

pronouns in the dative, e.g. "to me"	ma-Ga	ma-Gan
the possessive suffix for <i>sender</i> ("you", plural, informal)	-Nar, -ner, -nör	-Ndar, -Nder
The formation of Future Tense	–	-baq / bek-, -paq / pek-, -maq / mek
endings in the 3rd person plural, present tense	-(she)t, as in <i>bari-shat</i> ( <i>they go</i> )	-di, -dī
endings in the 3rd person plural, past tense	-d-ıştı, -d-ishti	-di, -dī

**Note:** The rather odd Kyrgyz formation *bari-shat* "they go" apparently results from the superposition of the mutual mood marker *-sh-* and a posterior vowel metathesis: *bari-sh-tür* > *bari-sh-ti* > *bari-sha-t*.

### The lexis of Kyrgyz and Kazakh

*Kyrgyz seems to be a rather archaic language with a minimum number of lexical borrowings*, which clearly sets it apart from Kimak that includes a number of Oghuz innovations and Perso-Arabic loanwords (see below).

Speakers of both Kazakh and Kyrgyz usually report good mutual intelligibility and sometimes state that they are *bir tuGan* "of one kin". The differences in Swadesh-215 seem to be *very small*, no more than 8%, and in some cases these are just minor inconsistencies in dictionaries. Only the following clear-cut mismatches were found in the original Swadesh-200:

	Kyrgyz	Kazakh
leg	<i>but</i> (as in Altay), also <i>ayaq</i> "foot"	<i>ayaq</i>
big	<i>choN</i> , apparently from Altay <i>Ja:N</i> . Also <i>ulu</i> : "great"	<i>ülken</i>
what	usually <i>emne</i> , also frequently <i>ne</i>	<i>ne</i>

that	tigi	<i>anau, sonau</i>
sniff, smell	usually <i>jito:</i> , but more literary or formal <i>iisko:</i>	<i>iiskeu</i>
sing	<i>i:rdö:</i> , also <i>jirlau</i> (?)	<i>zhirlau</i>
wet	<i>nim, nimdu:</i> (< Perso-Arabic <i>nam</i> "moisture")	<i>ilgal</i>
to swell	<i>köbü:</i> , <i>shishü:</i>	<i>isip-kebu, isinu</i>
sharp	<i>kurch</i> , also <i>ötkür</i>	<i>ötkir</i>
thin	<i>ichke, jukê</i>	<i>zhiñishke;</i> <i>zhûqê</i> "fine, thin work"
to burn	<i>küyü:</i> , also <i>janu:</i>	<i>zhanu</i>
to hear	<i>ugu, eshitu</i> (probably outdated or dialectical)	<i>estu</i>
correct	<i>tu:ra</i> , sometimes <i>durus</i> "decent, right"	<i>dûris</i>
feather	<i>tal jünü, jün</i>	<i>qawirsin</i>
rain	<i>JamGir</i> (a normal Great-Steppe variant)	<i>zhaNbir</i> (probably changed because of the Oghuz * <i>yaGmir</i> ); <i>zhauin</i>
tree	<i>JiGach</i> (looks like a local Tian-Shan development, also found in Karakhanid <i>yiGach</i> )	<i>aGash</i>
wipe	<i>aarchu</i> , also <i>sürtü:</i>	<i>sürtü</i>

Among the local Kyrgyz-Kazakh isolexemes, shared by Kyrgyz and Kazakh but apparently absent from other languages (except from the affiliated branch of Chagatai where they must have appeared from Proto-Kyrgyz-Kazakh), the following examples could be found:

Kyrgyz *kiiyö*, Kazakh *kiiyeu* "husband";

Kyrgyz *chöp*, Kazakh *shöp*, Uyghur *chöp*, "grass";

Kyrgyz *sogu.*, Kazakh *soGu* "blow (of wind) (originally: strike)"; Kyrgyz *qachiq*, Kazakh *qashiq* "far away" (from *kach-* "to run away");

Kyrgyz *soru.*, Kazakh *soru* "suck" also exist in Altay-Khakas and/or Uzbek-Uyghur but seem to be absent or not typical in Tatar-Bashkir;

Kyrgyz *özön*, Kazakh *özön* "river", typical in this meaning only of Kyrgyz-Kazakh, though is also known in Kumyk, Tatar, Salar, Altay, etc as "brook", "stream" and Crimean Tatar "river" (which may be an independent semantic mutation);

Also, cf. the phonological similarities in

Kyrgyz *jumurtqa*, Kazakh *zhûmurtqa* "egg";

Kyrgyz *jalbîraq*, Kazakh *zhapîrak* "leaf", which are rather unique among other Turkic (and presumably archaic).

### **The history and geography of Kazakh**

The Kazakh Khanate was founded in 1456-1465 by Janybek (Zhany-bek) Khan and Kerey Khan in the Jetti-Su area (in the southeastern part of present-day Kazakhstan), following a successful rebellion against the Uzbek Ulus and its Abu'l-Khayr Khan. [These events were described by Mukhammed Khaydar in *Tarih-i-Rashidi*]. The early years of the Kazakh Khanate were marked by the struggle against the Uzbek leader Muhammad Shaybani, who was defeated in 1470.

Consequently, the Jetti-Su (Zhetysu) ("The Seven Waters") area north of Almaty and especially *the area of the Chu river*, can be regarded as the Kazakh Urheimat, where the Kazakh Khanate was first founded and where the Kazakhs began their expansion to the Great Steppe in the north.

On the other hand, the Chu River, that now runs along the Kazakh-Kyrgyz border from the present-day territory of Kyrgyzstan, is often seen as a traditional Kyrgyz habitat just as well. Actually, this is where Bishkek, the capital of Kyrgyzstan, is located. Almaty, the largest city of

Kazakhstan, is only 200 km (120 miles) away from Bishkek across the Zaili (=from Russian *Za-Ili-yskiy* "Trans-Ilian, behind the Ili River") Alatau Ridge, so both settlements are situated at the foot of the Tian Shan Mountains nearly in the same area. Consequently, the geographic and historical connection between the Kyrgyz and Kazakh ethnicities becomes quite evident.

### **The dialectal differentiation in Kazakh**

There are at least two major dialectal groups within the Kyrgyz language: the Northern and Southern dialects. This dialectal differentiation in Kyrgyz marks it as a slightly "older language" than Kazakh, which is much more dialectically uniform. Indeed, despite the large territory it occupies, Kazakh is often reported to have no dialects at all, especially in popular, nonscientific sources. However, this is not entirely true. The Western Kazakh dialect may differ (or may have differed in the past before the mass Russification and the TV standardization began) from the Eastern one in several ways, including such features as the Western /zh/ : Eastern /j/ pronunciation, the usage of *-zhaq* / *zhek* for the future tense, etc.

Moreover, certain minority dialect-languages in Astrakhan (along the Volga) can presently be viewed as nothing but westernmost dialects of Kazakh, since they share 98% of mutual intelligibility with it, e.g. the so called Karagash Nogai language (not to confuse with Nogai Proper on the Caspian Sea) and Karakalpak.

In any case, the weaker dialectal differentiation in Kazakh as compared to Kyrgyz marks it as a little "younger" language that must have been spreading north from the area of stronger dialectal differentiation, such as the foot of the Tian Shan Mountains near Kyrgyzstan but was affected by the dialect of Nogai clans in the Great Steppe south of the Urals.

### **Alternative taxonomic hypotheses**

The placement of Kyrgyz within the same subgroup as the Altay Turkic languages was popularized by the famous Baskakov's classification, which became a generally-accepted standard in the Soviet-Russian Turkology [see Baskakov, N.A. *Klassifikatsiya tyurkskikh yazykov v svyazi s istoricheskoy periodizatsiyey ikh razvitiya i formirovaniya* (*The classification of Turkic languages as connected to the historical periodization of their formation and development*), Moscow (1952)]. However, judging by his later works from the 1960's to 1988, it turned out that there was no or little specific argumentation for this taxonomic decision. Generally speaking, Baskakov's classification was based on phonological and grammatical features, and some personal intuition, without any vocabulary comparison.

### **Conclusions:**

The close relatedness between Kazakh and Kyrgyz is hardly deniable. In fact, they are so lexically close (92%, Swadesh-215) that under certain simplifying circumstances they could even be viewed as very distant dialects or variants of each other, however, the notable discrepancy in phonology and grammar marks them as distinct languages.

We can now draw several conclusions concerning the early Kazakh history. Based on (1) the weaker dialectal differentiation in Kazakh as compared to Kyrgyz; (2) the presence of notable Nogai phonological features; (3) the geographical proximity of Kazakh to the languages of the Golden Horde, particularly Nogai; (4) its original location along the Chu River, near the present-day Kyrgyzstan border, Kazakh can be viewed as a historically recent 14th-16th century expansion of Kyrgyz-related tribes from the Tian-Shan Mountains into the northern steppeland. Because of the expansion over the large territory of the Kazakhstan steppe, the early Kazakh tribes must have made contact with various languages and dialects of the Golden Horde, specifically the early Nogai and other Kimak-related dialects along the Volga and the Ural (Yaik / Jaik) River. This contact may have resulted in the formation of a "Nogacized" form of the medieval Kyrgyz, which finally led to the emergence of the present-day Kazakh and Karakalpak languages.

### **Altay-Kyrgyz isolexemes**

Besides the close proximity between Kazakh and Kyrgyz, there also exist several Altay-Kyrgyz isolexemes, which make the Kyrgyz relationship with Kazakh less apparent:

#### **Altay and Kyrgyz lexis and phonology**

In basic vocabulary, both Altay and Kyrgyz share a number of isolexemes:

- (1) Altay *jaan*, Kyrgyz *choN*, and Uyghur *chong* "big";
- (2) Altay *kurch*, Kyrgyz *kurch* "sharp (as of a knife)";
- (3) Altay *moko*, Kyrgyz *mokok* "dull (as of a knife)", also cf. Tuvan *mugur*, probably from Mongolian;

(4) Altay *d'ün*, Kyrgyz *jün*, Khakas *chüg* "feather" as opposed to Kazakh *qawırsın*;

(5) Altay *sok*, *sogor*, Kyrgyz *sogu*.; Kazakh *soGu* "to blow (as of wind) (literally "to strike");

(6) Altay *uk*, Kyrgyz *ugu*: "to hear"; also found in Khakas, Uyghur, Kazakh as "to understand", though this word is more typical of the Altay dialects than any other languages. The word may be related to the Mongolian *uqa-/uxa-* "to understand" [see Sevortyan's dictionary (1974)];

(7) Altay *küyer*, Kyrgyz *küyü*: "to burn (intr.)", also attested in Khakas, Tuvan;

Among examples of lesser importance, one can also note:

(8) Altay *sler*, Kyrgyz *siler*, not to confuse with *sizder* "you (plural)", cf. a similar but not identical Kazakh secondary formations *sen-der*, *siz-der*. The *siler* isolexeme is obviously not exclusive to Kyrgyz-Altay, but is widely used in Altay-Sayan, Uyghyr as well as probably in some other Turkic languages east of the Tian Shan;

(9) Altay *bul*, Kyrgyz *bul*, Kazakh *bül*, and also Bashkir *bil* "this", instead of the apparently more archaic *\*bu* (and despite the alleged Starling's external etymologies, where the Altaic words for "body" seem to be used). However, this particular phonological shape was picked up much earlier, before the separation of Kazakh and is rather archaic;

Moreover, note the following phonological similarities:

(1) Altay *üren*, Khakas *üren*, Kyrgyz *ürön* "seed", as opposed to Kazakh *ûriq*, Uzbek *uruG*, Uyghur *uruq*;

(2) Altay *sö:q*, Khakas *sö:q*, Tuvan *sö:q*, Kyrgyz *sö:q* "bone", as opposed to Kazakh *süyeq*, Uzbek *suyoq*, Tatar *söyaq*;

(3) Altay *o:s*, Khakas *a:s*, Tuvan *a:s*, Kyrgyz *o:z* "mouth", as opposed to Kazakh *awız*, Tatar *avız*;

In other words, the typical Altay-Sayan phonological contraction that we have discussed earlier in the chapter dedicated to Altay-Sayan is also present in Kyrgyz, at least to some extent.

## Kyrgyz history

One of the most dramatic historical periods in the history of the Kazakh nation was marked by the long-lasting struggle (1723- 1758) against the Dzungarian Khanate that ruled over East Turkistan and West Mongolia in the 18th century. This severe and brutal conflict finally forced the Kazaks to seek alliance with the Russian Empire in 1731.

It is assumed herein that this period could also be marked by *the presumable Altay-Kyrgyz migrations*, which might have brought Altay Turkic to the Tian Shan Mountains where it intermingled with the local Kyrgyz language. This tentative hypothesis is corroborated by the fact that some similar Altay—Tian-Shan migrations are mentioned in the *Manas*, the Kyrgyz epic. Some corroboration may also be reflected in the ethnonymic conflation between the Altay-kizhi people (=Standard Altay speakers living in the Altai) and the Oirots (=Dzungarians of Mongolic origin near the Mongolian Altai), since the Altay-kizhi retained the name of Oirots or Oirats well into the Soviet era. This conflation suggests that some the Altay-kizhi could have become part of the Oirat army and participated in the invasion of the Tian Shan.

It is also known from historical records that the Kyrgyz people had been pushed by the Oirat invasion into the Ferghana valley [*The Great Russian Encyclopedia* (2005)]. Moreover, some of the Mongolic Oirats, known as *Sart-Kalmaks*, survived the downfall of the Dzungarian Khanate (1755-58) and became part of the Kyrgyz tribes staying near Lake Issyk-Kul.

If this conjecture is true, all the changes in Kyrgyz that differentiate it from Kazakh and make it similar to Altay must be relatively recent and acquired just a few centuries ago.

### **Kyrgyz geography**

The present-day mountain habitat of the Kyrgyz people in the Tian Shan appears to be *a typical isolated refugium* formed after several military invasions from the Kazakhstan steppe and Taklamakan desert, such as the Mongolian invasion (c. 1220-1450), and the Dzungarian invasion (c. 1720-1750's). This predicts an early Kyrgyz presence along the northern part of the Silk Road *in the Jeti-Su (Zhetisu) area and the Ili Valley* during the early Middle Ages. This earlier and more eastern habitat at the foothills of the Tian Shan was later superseded by the arrival of Kara-Khidans, Mongols, Dzungarians, and other invaders, making the Kyrgyz migrate closer to Lake Issyk-Kul in the Tian Shan.

### **Conclusion:**

Since many or most of the Altay-Kyrgyz isolexemes are equally found in Khakas and sometimes even Tuvan, and (1a) Altay has been shown above to belong to the Altay-Sayan taxon, on one hand, and (1b) Kyrgyz has been shown above to be closely related to Kazakh, on the



other hand, and (2) few of these words are found in the closely related Kazakh language, we may conclude that most of these unexpected Altay-Kyrgyz isoglosses are *late borrowings brought into Kyrgyz from Altay Turkic* somewhere between the 1500-1900's, that is already after the separation of Proto-Kazakh.

The most likely historical event that occurred in this geographic region during that historical period was *the Dzungarian invasion of the 18th century*. Therefore, we may assume that there existed an 18th century's military migration from the Altai to the Tian Shan Mountains, which brought these originally Altay lexemes into Kyrgyz, making the Kyrgyz language presently look more similar to Altay Turkic than it actually may be.

In any case, we must infer from the lexical evidence above that Kyrgyz is still more closely related to Kazakh than to any other Turkic language, whereas the Altay-Kyrgyz shared features must result from a secondary interaction between Altay and Kyrgyz.

### **Chagatai looks like Karakhanid affected by Kyrgyz**

The Chagatai subtaxon includes medieval Chagatai, modern Uzbek, Uyghur and their dialectal variations.

### **The Chagatai subtaxon**

First of all, note that with just **86%** of lexical proximity in Swadesh-215 (obvious borrowings excluded), *the Uyghur and Uzbek languages (and their internal dialects) must be as close to each other as Turkish and Azeri*, which is the common example of closely related languages in the Turkic group and outside of it.

Both languages received their respective names only in the 1920's, being known as Chagatai, Sart or Türki for most of the time before that. The **Chagatai** subtaxon is often known as *Karluk* in Baskakov's classification and those of his followers. However, as we have explained above, the exact origins and linguistic affiliation of Karluks is very obscure, and it is far from clear what relation the early Chagatai people bore to the Karluk tribes. Moreover, this kind of misplacement of ethnonymic stress seems to make the Chagatai language and its well-known relatedness to Uzbek and Uyghur unjustly forgotten, which may make one wonder what kind of Turkic language Chagatai possibly was. For these reasons, the name "Karluk" for this taxon seems to be out-of-place and should probably be replaced with *Chagatai*.

### **Chagatai-Uzbek-Uyghur geography**

Just as the neighboring Kyrgyz, the Chagatai-Uzbek-Uyghur languages originally occupied mountain territories along the Tian Shan range as well as some of the suitable oases along the edges of nearby deserts.

**Note:** The Tian Shan is one of the longest mountain ranges in Central Asia forming part of the natural barrier between the Great Steppe in the north and the Taklamakan desert in the south. It merges with the Pamirs in the west and it is separated from the Altai by the Dzungarian Plane in the east.



A topographic map of the Tian Shan Mountains [topomapper.com (2011)]

### **Chagatai-Uzbek-Uyghur history**

The Chagatai Ulus was a Turko-Mongol Khanate inherited by Chagatai Khan (1183-1241), the second son of Genghis Khan (1162-1227), but ruled by his successors. The true founder of the Chagatai Ulus was Alghu, the grandson of Chagatai, who in 1261 established control over most of its territory but died in 1266.



Chagatai Khanate [en.wikipedia.org (2011)]

Giovanni da Pian del Carpine, who was passing through the Chagatay Ulus north of Tian Shan Mountains in **1245**, described some scenes of great devastation in the nearby western areas left after the war with the Mongols:

*Moreouer, out of the land of the Kangittæ [= probably, the land of Kangly located near the Ustyurt Plateau or nearby area], we entered into the country of the **Bisermini** [= apparently, a vague alias for Turkic-speaking Muslims, cf. dialectal Russian *basurmany* from *musulmany* "Muslims"], who speake the language of Comania [= by Cumania the author meant the vast land between the Kievan Rus in the west and the Volga River in the east, where Cuman-Polovtsian, or (Old) Kypchak, was spoken], but obserue the law of the*

*Saracens* [= Islam, Sharia]. *In this countrey we found innumerable cities with castles ruined, and many towns left desolate. The lord of this country was called Soldan Alti, who with al his progenie, was destroyed by the Tartars* [= the Mongols, Tataro-Mongols, Turko-Mongols, the Tatar tribes directed by the Mongols]. *This countrey hath most huge mountains* [= apparently, the Tian Shan]. *On the South side it hath Ierusalem and Baldach* [= Baghdad], *and all the whole countrey of the Saracens* [=Arabs, Muslims]. *In the next territories adioyning doe inhabite two carnall brothers dukes of the Tartars* [= Mongols], *namely, Burin and Cadan, the sonnes of Thyaday* [= Chagatai], *who was the sonne of Chingis Can.*

[Frier Iohn de Plano Carpini, *The long and wonderful voyage of Frier Iohn de Plano Carpini*, (1245-46)]

Political strife in the Chagatai Ulus never ceased since the days of its formation. In 1346, a tribal chief Qazag-Khan from the Mongolic tribe of Qaraunas in Afghanistan and eastern Persia [Babur noted that they still spoke Mongolian in the late 15th century] killed the Chagatai Khan-Qazan during a revolt. Qazan's death marked the end of an effective Chagatayid rule over Transoxiana. As a result, the administration of the region fell into the hands of the local chieftains of Turkic and Mongolic origin. Using the disintegration, Janibeg Khan, the ruler of the Golden Horde from 1342 to 1357, asserted Jochid dominance over the Chagatai Khanate.

**Note:** It is believed that Janibeg's army had catapulted infected corpses into the Crimean port city of Kaffa (1343) in an attempt to use the plague to weaken the defenders. Infected Genoese sailors subsequently sailed from Kaffa to Genoa, introducing the Black Death into Europe.

However, the Chagatayids expelled Janibeg Khan's administrators after his assassination in 1357. By 1363, the control of Transoxiana was contested by two tribal leaders, Amir Husayn (the grandson of Qazaghan) and the famous Timur, or Tamerlane. Timur [from Turkic *temir* "iron"] eventually defeated Amir Husayn and took control of the state.

As a legacy of the severe devastation caused by the Mongol invasion and the ensuing feudal turmoil, the Karakhanid language of the Tarim Basin lost its political dominance and cultural significance in the region. It is conjectured herein that the desolation of towns, the spread of deadly disease, the subsequent intervention of the Golden Horde and the resulting continual movement of large armies, as well as the later conquest of the Golden Horde territories by powerful Chagatai leader Timur (Tamerlane) resulted in *supplanting of the Karakhanid language by an unknown Great-Steppe dialect situated along the northern ridges of the Tian Shan Mountains, such as an early Kyrgyz or Karluk.*

Consequently, the early Chagatai language emerging during that period, was essentially a mixed dialect mostly based on the Kyrgyz grammar but with the Karakhanid phonology.

### **Chagatai-Uzbek-Uyghur phonology**

By taking a closer look at the actual lexical and phonological differences (see the table below), we may conclude that Uzbek and Uyghur phonology bears certain similarities to Karakhanid, e.g.:

(1) an innovative */\*S-/ > /y-/* mutation, just like in Orkhon-Karakhanid, e.g. Uzbek, Uyghur, Karakhanid *yol* "way" as opposed to Kyrgyz *jol*, Kazakh *zhol*; Uzbek *yurak*, Uyghur, Karakhanid *yürek* "heart" as opposed to Kyrgyz *jürek*, Kazakh *zhürek*;

(2) the retention of the nasal */-N-/* as in Karakhanid, cf. Karakhanid *müNüz*, Uzbek *mugiz*, Uyghur *müNgüz* "horn"; Karakhanid *süNük*, Uyghur *söNäk* (but Uzbek *suyak*), as oppose to Kyrgyz *sö:k*, Kazakh *süyek* "bone";

(3) the intervocalic or final **uvular** or **velar** */-G-/*, */-G/*, cf. Karakhanid *taG*, Uzbek *tôG* (mountain), Uyghur *taG*; Karakhanid *baGir*, Uyghur *beGir* "liver". By contrast, the languages of the Great Steppe all have */-w-/* and */-w/* in this case;

(4) the initial */b-/* **instead of** */m-/* just as in Karakhanid, cf. Karakhanid *boyun*, *boyin*, Uzbek *bûyin*, Uyghur *boyin* "neck", as opposed to Kyrgyz *moyun*, Kazakh *moyin*;

(5) the retention of the final */-vq-/* in certain words, such as in Karakhanid *yuvqa*, Uzbek *yupka*, Uyghur *yupqa* "thin", as opposed to Kyrgyz *Juka*;

(6) the lenition of the "heavy" */-d-/*, */-t-/* into the "lighter" */-l-/*, which provides Uzbek-Uyghur with a more lenitioned, more simplified and more western pronunciation as in Uzbek *-lar*, Uyghur *-lar*, *-lêr*, as opposed to Kyrgyz *-lar*, *-ler*, *-lor*, *-lör*, *-dar*, *der*, *-dor*, *dör*, *-tar*, *-ter*, *-tor*, *-tör* with its heavy, fortified consonants and some similar fortition in other languages of the eastern part of the Great Steppe.

On the other hand, the Great-Steppe phonological influence in general and the Kyrgyz influence in particular is also quite evident, cf.

(1) the innovative metathesis in Uzbek *yamGir*, Uyghur *yamGur* as in Tatar *yaNgir*, Bashkir *yamgîr*, Nogai *yamGîr*, Kyrgyz *jamgîr* and other languages of the Great-Steppe, instead of the Old Uyghur *yaG-mur* from *\*jaG-* "to fall, to rain" and *\*mur*, the typical Proto-Altaic word for "water";

(3) Uzbek *mûgiz*, *muguz*, Uyghur *müNgüz*, which is similar to the Kazan Tatar *mögez*, Bashkir *mögöđ*, instead of Karakhanid *müNüz*, Old Uyghur *müyüz*;

(4) Uzbek *sovuk*, which is similar to the Kazan Tatar *süwiq*, Bashkir *hüwiq*, Nogai *suwiq* instead of the Karakhanid *suGiq*, though it is also partly retained in Uyghur *soGaq*;

(5) Uzbek *yaproq* from Proto-Kimak *\*yapraq* instead of the longer *yapurgak* in Karakhanid, though the Old-Uyghur-Karakhanid pronunciation is also partly retained in modern Uyghur *yapurmaq*;

The table below lists some of the *phonologically dissimilar words in Turkic languages of Central Asia*. Note that Uzbek, Uyghur and Karakhanid are mostly colored dark red, marking their apparent lexical and phonological relatedness of Uzbek-Uyghur to Karakhanid, with just a few Kimak-Kypchak-Tatar borrowings in Uzbek.

A List of Phonologically Dissimilar Basic Words in Central Asian Turkic Languages							
	Turkmen	KazanTatar Bashkir SibirTatar CrimeanTatar Nogai	Kyrgyz, Kazakh	Uzbek	Uyghur	Salar	Karakhanid
not (adj, nouns)	däl	KT. tügel	emes	emas	emes	emes	ärmäs; ämäs (rare); täkül (cited only as Oghuz by MaK)
horn	buynuz; shax	KT. mögez;	Kg. müyüz;	muguz; shox	müNgüz	moNüz	müNüz, muNuz

		B. mögöð	Kz. müyiz;				
bone	süNk	KT. süyäk; B. höyäk N. süyek;	Kg. söök; Kz. süyek;	suyak	süNäk	senix	süNük
cold	sowuk	KT. siwïk, siuq ST. siuq B. hiwiq	Kg. suuk; Kg. suïq;	sovuk	soGaq	sox	soGiç
liver	baGir	KT. bawïr; B. bauïr	Kg. boor;Kz. bawïr;	zhigar	beGir	paGir	baGir
mouth	aGiz	KT. awïz; B. auïð	Kg. ooz;Kz. awïz;	oGiz	eGiz	aGiz	aGiz
mountain	daG	KT. tau; B. tau	Kg. too; Kz. tau;	toG	taG	taG	taG
neck	boyun	KT. muyïn; B. muyïn; N. moyïn	Kg. moyun;Kz. moyïn;	bûyin	boyun	poyni, puyni	boyin
round	öwre	KT. yomri;	Kg. Jumuru	yumaloq	yumlaq, yumilaq		yumGaq

		B. yomoro					
rain	yaGışh, yagmir	KT. yaNgir; B. yamGir N. yamGir	Kg. Jamgir;Kz. zhaNbir;	yomGir	yamGur	yaGmur	yaGmur
small	kichi	KT. keche; B. kese;	Kg. kichine- key;Kz.kishken-tay;	kichkina; kichik	kichik	kichi, kiJi	kichik
sleep	u:qla-	KT. yokla-; B. yoqla-; N. uyqla-; CT. yuxla-	Kg. uktoo, uyku;;Kz.üyıqtau	uxla-	uxla-	uxla-	uđi-
leaf	yapraG	KT. yapraq; B. yafraq;	Kg. Jalbıraq;Kg.zhapıraq;	(yaproq); barg	yopurmaq	yärfix, yaRfax	yapurGaq
dry	Gurı	KT. korı; B. qoro	Kg. qurGaq;Kz. qırGaq	quruq	quruq	quru, qurı	quruG
home	öy	KT. öy; B. üy;	Kg. üy;Kz. üy	uy	öy	oy	ev, äw



seed	toxum	KT. orliq B. orlok	Kg. üron;Kz. üriq	uruG	uruq	ashlix	uruG
bite	dishle-	KT. teshlê- B. teshlê-	Kg. tishte-;Kz. tiste-	tishla-	chishlä-	chishli-	tishla-
earth	topraG	KT. tufrak B. tupraq	Kg. topuraq;Kz. topıraq	tuproq	topa	torix, torax	tubra:q
tree	aGach	KT. aGach B. aGas	Kg. Jigach;Kz. aGash	yoGoche"wood"; daraxt	därax	ta:l	yıGach
grass	ot	KT. ülên B. ülên	Kg. ot; chöp;Kz. ot, shöp	ut	chöp	chöp	ot
thin	inche	KT. nechkê B. nêðek	Kg. ichke;Kz. zhiNishke	iNichka	inchikä	läshgi	yinchkä
thin (2)	yuGa, 'uka	KT. yuka	Kg. Juka	yupqa	yuqqa	yoxba	yuvqa
eat	iy-	KT. asha- B. asha- N. asha-	Kg. Je-;Kz. zhe-	ye-	yä-	yï-	yä-
belly	Garin	KT. korsak B. qorhaq	Kg. qarın;Kz. qarın;qûrsaq	qorin	qo(r)saq	xusax	qarın

### Chagatai-Uzbek-Uyghur grammar

However, *the Uzbek-Uyghur grammar usually lacks the most essential Orkhon-Karakhanid features* ( and they may only be occasionally present in Chagatai), namely:

(1) the lack of the archaic copula *er-/är-* (see below) and its mutation to *e-* in Uzbek *e-mes*, *e-dim* just like in other languages of the Great Steppe; neither is there any notable usage of *oftägül* which was known in Old Uyghur;

(2) the lack of the typical Karakhanid usage of the 3rd pers. singular pronoun *ol* as a copula (see below), e.g. *ul mâniN oGlim ol*, literally "he (is) my son-he". The *ol*-copula mutated to zero in modern Uzbek-Uyghur languages;

(3) the absence of the Future Tense with *-Gay*, *-gey* (see below) in Uzbek-Uyghur known in Karakhanid, Old Uyghur and other representatives of the Southern branch, though it sometimes may be retained in written Chagatai as *-Ge*;

(4) the absence of the archaic instrumental case ending *-(n)in*, that was originally present in Karakhanid, Old Uyghur and other early branches of Turkic Proper;

(5) the lack of the archaic directional case ending *-Garu* known from Old Uyghur and other representatives of the Southern branch;

(6) no persistent usage of *-mish-* (replaced by *-Gan-* as in other languages of the Great Steppe), though *-mish-* is still sporadically present in Chagatai and Uzbek dialects.

The situation with the *-mish-* seems to be more complex than it may initially seem, since *-mish-* can be used quite actively in modern Uzbek (as an example consider the song provided as an example in *The Turkic languages in a Nutshell*, ), but seems to be absent from the published grammars of the "literary" Uzbek. That may imply that the grammar of Standard Literary Uzbek is the same kind of science fiction as those of Standard Khakas, Altay, Evenk, Nenets, etc.

**Note:** The creation of "literary" local languages (sometimes renamed herein as "standard" in English), was part of the general paradigm in the postwar Soviet Union. Since it was quite difficult or even impossible to conduct specific research for each and every local dialect and separate all the dialects from all the languages, certain simplifications had to be made with some major dialect getting clustered into a single category and the local particularities being ignored and forgotten. In some cases, this procedure could even lead to the loss of the intelligibility with the proclaimed literary standard or a virtual loss of the vernacular.

As a matter of fact, the most typical grammatical features of modern Uzbek and Uyghur clearly point to the languages of the Great Steppe, particularly Kyrgyz and Kazakh. Consider the following Uzbek-Uyghur morphemes:

(1) The typically Great-Steppe verbal ending **-di / -dī / -ti / -tī in the 3rd person singular in the present and future tense**, e.g. Uzbek *bor-ap-ti* "he is going", *bar-a-di* "he will go", Uyghur *bar-i-du* "he'll go", *yaz-i-du* "s/he, they (will) write", cf. Kyrgyz *bar-a-t* "he will go", Kazakh *bar-a-dī* "he is going".

(2) The usual Great-Steppe verbal ending **-d-ik in the 1st pers. plural Past Tense**, cf. Uzbek *bor-d-ik* "we went", *kel-d-ik* "we came", Uyghur *yaz-d-uq* "we wrote" as in Kyrgyz *bar-d-ik*, *kel-d-ik*, even though it seems to be used interchangeably with the Karakhanid *-dimiz > -divuz* in the Toshkent dialect of Uzbek, cf. *bar-d-uvuz* "we went", *kel-d-ivuz* "we came". The *-d-ik* type of suffix also seems to be occasionally attested in Karakhanid sources in relation to Oghuz, but it had never been original to the Orkhon-Karakhanid subtaxon.

(3) The typically Kyrgyz-Kazakh **-īb-man, -īp-tir Unexpected Past Tense** as in Uzbek *unut-ib-man* "so it turns out I forgot", Uzbek *kel-ip-ti* "so he really came", Uyghur *yez-īp-tu* "he(really) wrote", cf. Kyrgyz *al-īp-tir* "so it turns out he took it, he really took it", Kazakh *söyle-p-ti* "he seems to have said", *bar-īp-pin* "I might have gone".

(4) The **-yat-ir-man Present Continuos Tense** as in Uzbek *yaz-a-yat-ir-man* "I am writing", Tashkent Uzbek *bor-wot-ti* "he is working" (a contracted form), Uyghur *kir-i-wati-men* (a contracted form) "I'm coming in", cf. similar forms in Kazakh *bar-a-zhat-ir-min*, Kyrgyz *bar-a-jat-a-min* "I walk, I'm walking", Kyrgyz *oku-p-jat-a-min* "I'm reading". The original grammatical meaning was actually "I am lying doing something" which perhaps initially implied a leisurely, slow passage of time as if resting in a yurt. The *-a-* suffix here seems to be just a spoken contraction from the *-īp-* gerundial suffix, given that the latter is much more widely used in Kyrgyz and Kazakh in similar expressions.

(5) The typically Central-taxon **-Gan Perfect Tense** normally absent from the Southern taxon where Karakhanid and Old Uyghur belong, e.g. Uzbek *ishla-Gan-man* "I have worked", Modern Uyghur *yaz-Gan-män* "I have written", cf. Kazakh *ol kel-gen* "he has come", Kazakh *men kel-gen-min* "I have come", etc.

(6) The widely used **-a-man, -y-man, -e-men Habitual Present / Future Tense** instead of the *-r-* Aorist in Old Uyghur and Karakhanid, e.g. Uzbek *ishla-y-man* "I work; I will work", Uzbek *men bil-ma-y-man* "I don't know", Uyghur *kir-i-men* "I enter", cf. Kyrgyz *bar-a-min* "I will go", Kyrgyz *bil-be-y-min* "I don't know", Kazakh *bar-a-min* "I will go", Kazakh *bol-a-min* "I will be". The Aorist in Uzbek-Uyghur is now used only in the meaning of a potential or uncertain future, e.g. Uzbek *bar-ar-man* "I think I will go", Uyghur *kir-ir-men* "I might enter", Uyghur *tut-mas* "he might not catch (hold) it".

(7) The **-mak-chi-men Tense expressing wish or intention**, e.g. Uzbek *qil-moq-chi-man* "I'm going to do it", Uyghur *yaz-maq-chi-men* "I'd like to write", cf. Kyrgyz *yaz-mak-chi-min* "I want to write". The construction originally meant "I am the doer (for this)" > "I'm eager to do this"; and it does not seem to be attested in Karakhanid.

(8) The **-Gin / -Gin imperative** as opposed to *-Gil / -Gil /-qil /-qil* imperative in Karakhanid and Old Uyghur, e.g. Uzbek *oqi-gin* "You read!", Uyghur *yaz-Gin, yez-iN* "You write!", cf. Kyrgyz *bar-gin* "You go!", but Karakhanid *tur-gil* "Stand up!".

### Chagatai-Uzbek-Uyghur lexis

But to which subgroup within the Great Steppe taxon is Chagatai-Uzbek-Uyghur related most?

According to the lexicostatistical research (2012), there is about **83%** of average distance from Uzbek-Uyghur to Kyrgyz-Kazakh, about 78% to Tatar-Bashkir, and about 74% to Turkmen (with borrowings excluded), which *marks Kyrgyz-Kazakh as the most closely related subtaxon (outside Orkhon-Karakhanid which could not be counted lexicostatistically)*.

Uzbek-Uyghur and Kyrgyz-Kazakh seems to share a few presumably innovative isolexemes in Swadesh-215 that are apparently missing or rare in other subgroups, cf.

- (1) Uzbek *yi qilmoq*, Uyghur *yi qilmaq*, Kazakh *zhīGīlu*, Kyrgyz *zhīGīlu* "to fall";
- (2) Uzbek *dumaloq*, Uyghur *domlaq*, Kazakh *domalaq* "round (such as wheel, lake, table)";
- (3) Uyghur *chöp*, Kazakh *shöp*, Kyrgyz *chöp* "grass";
- (4) Uzbek *uqalamoq*, Uyghur *ugulumaq*, Kazakh *uqalau*, Kyrgyz *ukalo*: "to rub";

Moreover note certain Great-Steppe words with some wider distribution in the nearby languages:

(5) Uzbek *bu yerda*, Uyghur *bu yerde*, Kazakh *būl zherde*, Kyrgyz *bul zherde* "here", also at least in Altay *bu d'erde* and Turkmen *bu yerde* "here". This phrase, of course, is not necessarily originally Kyrgyz-Chagatai or even Great-Steppe; it may have formed at an earlier level or even independently in several Turkic subgroups with some posterior contact spreading (for instance, probably into Turkmen which often borrowed from Great-Steppe). Nevertheless, its usage in the Kyrgyz-Chagatai subgroup in the sense of "here" is quite typical.

(6) The verb *kil-* in its direct meaning of "to do" seems to be particularly common of Kyrgyz, Uzbek, Uyghur, Bashkir, however it is not limited only to these languages and is widely distributed in various meanings from Tuva to Turkey.

(7) Uzbek *tüshün-*, Uyghur *chüshen-*, Kazakh *tüsin-*, Kyrgyz *tüshün-*, Tatar *töshen-*, Karachay-Balkar *tüshün-*, Kumyk *tüshün-*, Turkmen *düşün-* has the meaning "to understand", for the most part, only in the above-listed languages, even though it may also be distributed in

other branches in similar meanings, e.g. Turkish, Gagauz and Azeri *düşün-* "to think", Nogai "to look into something, to study" and Kumyk "to guess", etc. [Verified with *Sevortyan's Dictionary*]. It seems that the meaning "to understand" was formed in the Great-Steppe subtaxon, whence it spread into Oghuz-Seljuk (or vice versa). The original meaning of this verb in the literal translation was "to fall oneself; to be fallen" from *\*tüsh-ün-* as if "I fall myself; I'm being fallen (into this)" as in the English idiom "it sinks in".

(8) Uzbek *tovush*, Uyghur *tawush*, Kazakh *dawıs*, Tatar *tawış*, Bashkir *tawış*, Karachay-Balkar *tawush* "voice", a Great-Steppe innovation.

(9) Uzbek *uy*, Uyghur *öy* "home, house", most Great-Steppe languages *\*iyy*.

These 4 words constitute merely 2% in Swadesh-215, so it is hard to make any claims concerning particular relatedness of Uzbek-Uyghur to Kyrgyz-Kazakh. However, the general trend in the analysis of the vocabulary described above is to exclude the Kimak subgroup from direct Chagatai predecessors.

That becomes even more evident if we take into consideration the closer geographic proximity between Kyrgyz-Kazakh and Chagatai-Uzbek-Uyghur, as opposed to Kimak tribes scattered somewhere near the Urals.

By the same token, there are no grounds to suggest that Proto-Kazakh could have affected Proto-Chagatai in a direct way, since we know from history that the formation of Chagatai must have occurred before the separation of Kazakh from Kyrgyz, which is corroborated by the lack of any Kazakh-exclusive isolexemes. Quite on the contrary, we have:

(1) Kyrgyz-Chagatai *\*yamGur* "rain", but Kazakh *zhaNbir*;

(2) Kyrgyz-Chagatai *\*qil-* "to do", but usually Kazakh *isteu*, *zhasau*;

Consequently, we should infer that *the Great-Steppe tribe that came in contact with Karakhanid in the 13th-14th century belonged to the Kyrgyz-Kazakh subgroup*, thus resulting in the formation of the early Chagatai, whereas the Kimaks or the early Kazakh tribes could not have played any significant role in this exchange. The tribal unity under consideration could be Karluk, but there is no direct linguistic evidence.

### **Conclusion:**

It all looks as if Proto-Chagatai were a language of newly-arrived Kyrgyz-related speakers who continued to build sentences in the way similar to modern Kyrgyz or Kazakh but adopted the Karakhanid-style pronunciation, e.g.

Proto-Uzbek-Uygur \**müNüz* cf. Karakhanid *müNüz*, instead of Kyrgyz *müyüz* "horn";

Proto-Uzbek-Uygur \**taG* cf. Karakhanid *taG*, instead of Proto-Kyrgyz-Kazak \**taw* "mountain"; Proto-Uzbek-Uygur \**aGiz* cf. Karakhanid *aGiz*, instead of Proto-Kyrgyz-Kazak \**awiz* "mouth";

Proto-Uzbek-Uygur \**boyun* cf. Karakhanid *boyun* or *boyin*, instead of Proto-Kyrgyz-Kazak \**moyun* "neck";

Proto-Uzbek-Uygur \**quruq* cf. Karakhanid *quruq*, instead of Proto-Kyrgyz-Kazak \**qurGaq* "dry";

Proto-Uzbek-Uygur \**ye-* cf. Karakhanid *ye-*, instead of Proto-Kyrgyz-Kazak \**je-* "to eat";

Proto-Uzbek-Uygur \**yupqa* cf. Karakhanid *yuvqa*, instead of Proto-Kyrgyz-Kazak \**juqa* "thin";

However, many Karakhanid words were replaced by their Great-Steppe and Proto-Kyrgyz-Kazak equivalents, such as \**üy* "home, house" instead of Karakhanid *äv*; often \**qorsaq* instead of Karakhanid \**qarın*; \**yamGur* "rain" with a metathesis instead of Karakhanid *yaGmur*, etc.

Consequently, we can see that the Chagatai-Uzbek-Uyghur languages seems to inherit the original Kyrgyz grammar and some of the vocabulary, but acquired superficial phonological similarity to Karakhanid. The retention of grammar and lexis is normally more fundamental than the changes in the phonology that can be achieved more easily. Therefore we may conclude that the original Karakhanid speech of the 10th-12th centuries has not survived in the Tian-Shan and Taklamakan being overrun during the complex turmoil and ethnic disorder of the 13th century's Mongol invasion by a new speech of the newcomers from the the northern foothills of the Tian-Shan Mountains who spoke a Kyrgyz-related dialect. (The only living direct descendant of Southern Karakhanid seems to be Khalaj, as shown below.).

A counter-argument that Karakhanid and Old Uyghur may be poorly attested and perhaps possess some of the grammatical features described in here as purely Great-Steppe is implausible, judging from the fact that these grammatical features are equally absent from Oghuz-Seljuk languages (the closest modern Karakhanid sibling), and still mostly belong to Proto-Kyrgyz-Kazakh.

Approximate glottochronological calculations suggest that the separation of Proto-Chagatai from Proto-Kyrgyz-Kazakh must have occurred at least a few centuries before the Mongol invasion, c. 1000 AD, so it is difficult to attribute Proto-Chagatai directly to the early Kyrgyz, rather it could have been a slightly different Kyrgyz-related dialect, possibly such as Karluk, though the linguistic affiliation of the latter remains unknown.

**Note:** The formation of such "mixed" languages is a typical adstratic phenomenon occurring at the boundary of two ethno-geographical areas, sometimes involving strong impact from a third or fourth superstratic component (in this case, Arabic and Persian). This interaction usually leads to remarkable, historically rapid changes in a language, and without a doubt deserves a separate detailed consideration elsewhere.

Additionally, Standard Literary Uzbek or its dialects could have picked up certain lexical and phonological elements from Kimak-Kypchak-Tatar languages, but that process must have been fairly recent, less significant and did not affect the basic vocabulary of Uzbek to the same extent.

The term *Karluk* should not be directly conflated with the dialects of Chagatai, Uyghur and Uzbek as in Baskakov's classification. The Karluks were an early Turkic clan confederacy of unknown dialectal affiliation that lived near the Tian Shan between the 8th and 12th centuries.

A suitable self-explanatory name for the Kyrgyz-Kazakh-Chagatai cluster could be *Tian-Shan*.

### **The Kimak subtaxon**

The *Kimak* subtaxon, sometimes also designated herein as *Kimak-Kypchak-Tatar*, includes at least the following languages and dialects:

(1) *the typical languages of the Golden Horde*, which include **Sibir Tatar, Bashkir, Kazan Tatar, Mishar Tatar, Nogai, Kумык, Northern Crimean Tatar, Lithuanian Karaim, Crimean Karaim;**

(2) **Baraba Tatar** (presumably separate);

(3) **Karachay-Balkar;**

The Kimak subtaxon does not include Kyrgyz or Kazakh.

Below, we will try to demonstrate that the above-mentioned Kimak languages indeed share common innovative features.

### **Kimak history and geography**

According to the work *Zayn-al-Akhbar* composed by Gardezi circa 1030, where he apparently cites the earlier writings by ibn Khordadbeh (820-912), there was the following legend about the Kimak origins:

Once upon a time, there were two sons left after the death of a leader of the Tatars. The younger son, named Shad, was envious of his elder brother, who was the heir to the kingdom, and attempted to kill him. Consequently, Shad had to run away with his slave concubine into the steppe near the Irtysh River, where they settled down in a yurt and lived happily for some time hunting squirrels and ermines. As a result, some of his Tatar relatives came over and joined them. These were the seven men named *Imi*, **Imak (Yamak, Kimak)**, *Tatar*, *Bayandur*, **Kipchak**, *Lanikaz*, and *Aj(a)lad*. All of them also settled down the Irtysh, and finally formed the seven tribes named after these forefathers. See [*Gosudarstvo kimakov IX-XI vv. po arabskim istochnikam (The Kimak State of the 9-11th century according to the Arab sources)*, Kumekov, B.E.; Alma-Ata (1972)]. Most authors writing on the subject [Kumekov (1972), Marquart (1920)] date this legendary period to about 700 AD, which is also confirmed glottochronologically herein.

For other details see [On the origins of the ethnonym Tatar](#).

By the time of Gardezi (c. 1030) and Mahmud al-Kashgari (c. 1070), the seven clans of the Kimak confederacy were well-established and described by several authors. Mahmud al-Kashgari cited an apt saying, "*The snake has seven heads*", referring to the Kimak clans. The Arab geographer [Al Idrisi](#) (1099-1165), who created his famous (though very convoluted by modern standards) map of the world known in Europe as *Tabula Rogeriana*, too mentions the existence of 16 Kimak towns apparently located in the upper Irtysh basin near Lake Zaysan [see figures below].

Therefore, the Proto-Kimak-Kypchak-Tatar tribes must have lived somewhere along the upper course of the Irtysh River, where they finally formed their own Kimak Kaganate.

The difference between the attested ethnonyms Kimak (Kimek) and Imak (Yemek) is poorly understood. We can hypothesize that the original name could have been preserved in the ethnonym *Kumyk* that may originate from a clan name, therefore the initial reading could be close to *\*Qimiq*, but this word was later misread or incorrectly recorded in the Arabic script with a different consonant.

In any case, the **Kimak (Kimek) Confederacy / Kaganate/ Khanate** was a prominent medieval Turkic state in the area of the middle and upper Irtysh River. It existed as the Kimak Kaganate from approximately 743 to 1050 AD, and as the Kimak Khanate until the Mongol conquest in the early 13th century (?). Even though the Kimaks were essentially nomadic, they had many cities mostly in the Irtysh basin, such as *Imakiya*, which was the summer seat of the Kimak kagan, and which is said to have markets and temples.

**Note:** the Arabic toponym *Imakiya* is probably a misspelling from *Kimakiya* /kee-mah-KEE-ya/ which is supposed to mean just "Kimak (City or Town)", for instance as in Arabic *al-arabi:ya*, *al-injli:ziya*, etc.



It can be inferred from the linguistic and ethnonymic evidence that during the 9th century CE, these Kimak tribes began to spread far away to the west. They were subsequently attested as (1) "Bashkirt" near the Southern Urals and the Volga River by Ibn-Fadlan in 921 and then as (2) "Tatar", "Bashkirt", "Kifchak", etc. by Mahmud al-Kashgari in 1073, as well as by other Arab authors. Consequently, they must have expanded as far as the Ural Mountains somewhere between the 750's-900's, or most likely, after the fall of the Göktürk-Uyghur Kaganate, that is after the 840's.

The period of the Kimak spread to the northwest is supported archaeologically: at some period between the 700-900 CE, there was a wave migrations into the Baraba Steppe that displaced the earlier Potchev culture in that area. The new culture was characterized by inhumations in burial mounds along with the horse, which is typically associated with the Turkic tribes. [*Arkheologija Zapadno-Sibirskoj ravniny (The Archaeology of the West Siberian Plane)*, Troitskaja, T.N., Novikov, A.V., Novosibirsk (2004), pp. 93-95].

Moreover, we may suppose that this migration must have proceeded along the northern and northeastern border of present-day Kazakhstan and Russia, because the Irtysh flows to the northwest providing a natural route for a travel in that direction.

The migration along the Irtysh towards the confluence of the Irtysh and Tobol is also corroborated by the existence of the Baraba Tatars along the middle course of Irtysh and the Sibir Tatars near the Tobol-Irtysh confluence. These ethnic groups share many common features both with each other and with the Bashkir and Kazan Tatars.

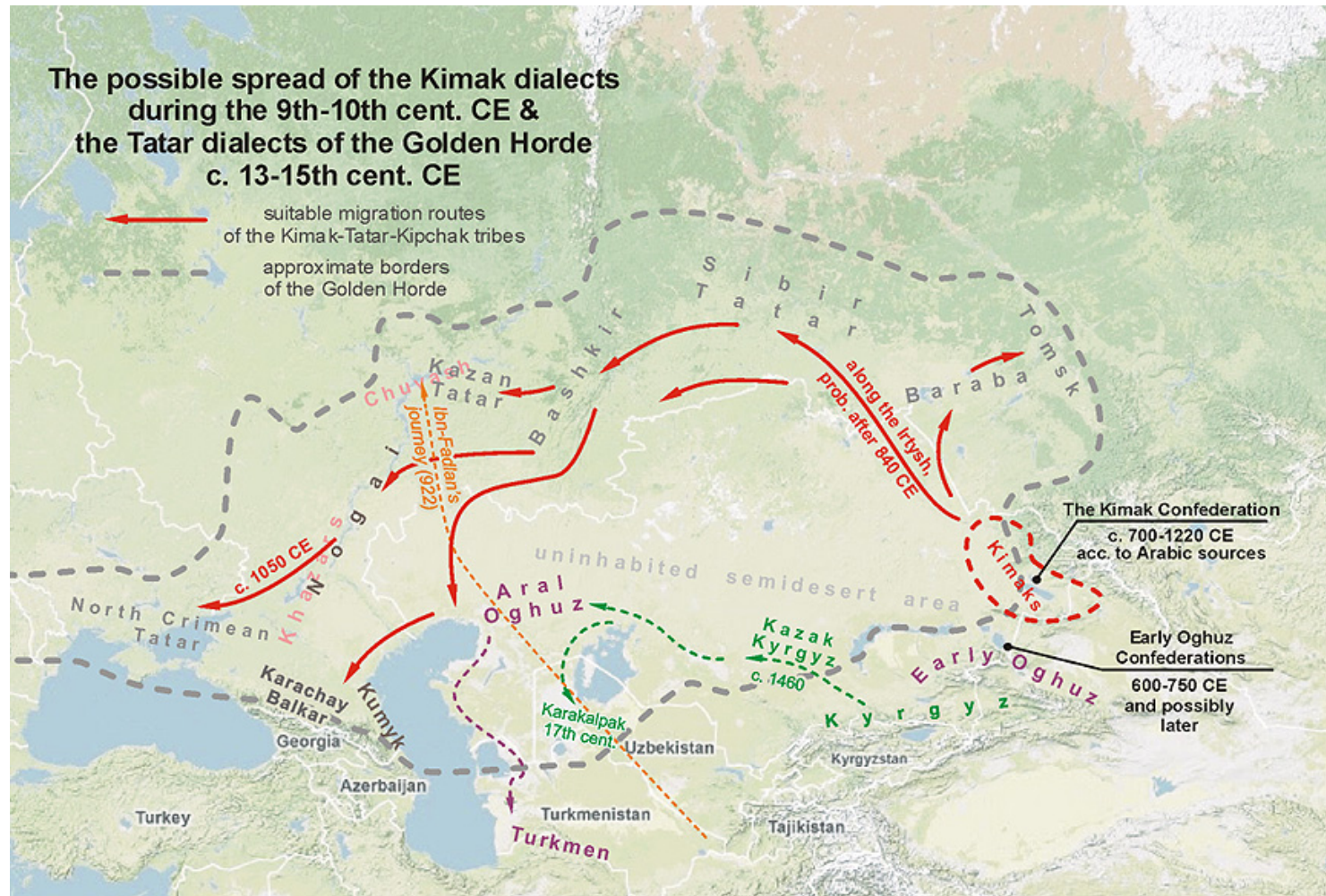
Otherwise, if the migrating Kimak tribes had turned west or southwest, they would have run into the Karluk and Kyrgyz territory in the south near the Tian Shan, mentioned by al-Idirisi and in other historical sources.

Also note that any direct migrations to the west across the central Kazakhstan are unlikely due to geographic difficulties, such as desert climate, highlands and the scarcity of water sources.

By following the Tobol and Yaik River, and/or traveling across the Southern Ural, the Kimak tribes must have crossed into Eastern Europe and formed the ancestors of the early Bashkirs and Tatars. Following the upper Kama, some of them must have reached the confluence of the Kama and Volga, where the Volga Bulgaria was located. These Kimak tribes must have become the precursors of what we presently know as the Kazan Tatar people.

The exact migration tracks of Proto-Northern-Crimean-Tatar, Proto-Karachay-Balkar, Proto-Nogai and Proto-Kumyk are harder to establish. At the time of their arrival to the Urals, all of these were almost linguistically indistinguishable, but they may well have belonged to different clans, so there still could be some genetic or political distinctions. Apparently, they split off from rest of the Kimak, Tatar and Bashkir tribes near the Southern Ural. Then, these tribes migrated southwest by following the Ural (Jaik) River first towards the Caspian Sea and the Caucasus Mountains, and finally as far as the Kievan Rus, where they soon became known as Kipchaks or Polovstians.

Most of the Kimak groups under consideration (or at least Kazan Tatar, Sibir Tatar, North Crimean Tatar, Caspian Nogai, etc) seem to have emerged as separate ethnicities with their own dialects only after the expansion and dissipation of the *Golden Horde* (1235-1502), and the formation of the localized post-Golden-Horde Khanates of the 16th century.



### The spread of the Kimak and Tatar dialects (2012)

It should perhaps be explained that the Golden Horde (cf. *ordu*, *orda* "army") is a historiographic name for the basically Kimak-Kypchak-Tatar Empire (1226-1502) established after the Mongol invasion of Rus and ruled by the nominal descendants of Genghis Khan. It was mostly known either as just *Orda* (in Russian sources) or as the (*Ulug*) *Ulus* "the (Big) Country" or by the name of its current ruler, such as *Ulus of Jochi* (in Turkic and Persian sources of that period). It was officially Islamized only in 1313.

The Golden Horde exacted taxes from Russians, Armenians, Georgians, Circassians, Alans, Crimean Greeks, Crimean Goths, and other subjugated peoples along its borders. The Golden Horde's capitals were (1) *Sarai-Batu* meaning "the Palace built by Batu Khan" and (2) just *Sarai* "the Palace", both of which were located along the Volga River and had many thousands of inhabitants. However, they were sacked, destroyed and dismantled after the fall of the empire.

The Golden Horde elite traced their descent from the Mongol clans and originally used the Middle Mongolian language as the main means of communication, however its most common population was apparently of Kimak-Kypchak-Tatar origin.

After the collapse of this powerful state by the end of 15th century, the newly-formed Kypchak-Tatar dialects and ethnic groups were for the most part vaguely known as "Tatars" to the Russians from the early 16th until the end of the 19th century. The word "Tatar" may still retain somewhat negative connotation in Russian and other languages affected by the expansion of the Golden Horde, including some European languages where *Tartar* became the synonym of "fierce" and "violent".

It is conjectured herein that nearly all the Turkic languages presently located on the territory of the former Golden Horde (Kazan Tatar, Mishar Tatar, Bashkir, Karachay-Balkar, Kumyk, Nogai, North Crimean Tatar, etc) are particularly close to each other to the extent of mutual intelligibility.

The Kimak languages share a number of distinct innovations in phonology, grammar and lexis. Some of these innovations are also shared with the Oghuz-Seljuk languages, an interesting phenomenon that deserves a separate description below. On the other hand, these Kimak innovations are mostly absent from Kyrgyz-Kazakh, that did not belong to Kimak or the Golden Horde, given that Kyrgyz was located far away in the Tian Shan Mountains, whereas Kazakh formed only after the middle of the 15th century when the Golden Horde no longer formally existed.

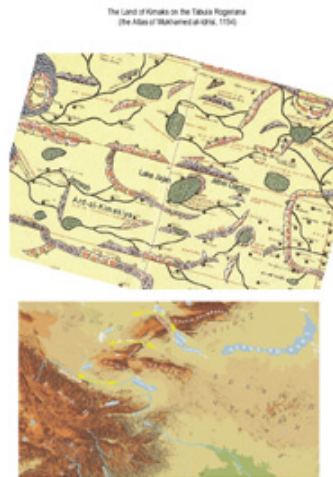
### **Kimaks on the map of al-Idrisi**

The location of the Kimak Confederacy was shown in the 12th century's atlas prepared by the Arab geographer Mukhamed al-Idrisi, known in Europe as the Tabula Rogeriana.

The Asian part of the map, which is extremely difficult to decipher, has been studied by several authors including Kumekov, B.E. in [*Strana kimakov po karte al-Idrisi (The land of the Kimaks according to the al-Idrisi's map)*// *Strany i narody vostoka*, vol.10, 1971, pp.194-198 (in Russian)].

Judging by phonetically garbled toponyms and the typical contractions and doubling, such as "Dardan", "Lalan", etc., the Asian part was probably based on some Chinese sources, assumingly on *hearsay evidence provided by medieval Silk Road merchants*. Consequently, the map is not grounded on astronomic measurements, and there is no such thing as scale or even orientation in it, so trying to link some of its features to modern geography can sometimes turn into a formidable task.

However, we may presume that the map features are supposed to match real-world geography to the extent that they would in a verbal account obtained from a medieval traveler, whereas the map toponyms are supposed to sound as if they were reinterpreted from the heavy Kimak-Tatar pronunciation into the medieval Chinese and then finally into al-Idrisi's Moroccan Arabic.



The Land of the Kimaks in the Tabula Rogeriana (clickable)

The map ends abruptly near Mongolia, where traveling in the Altai-Sayan Mountains was most likely impossible.

Apparently, B.E. Kumekov made an error by attributing *Lake Gagan* to Lake Alakol (Ala-Köl). It all becomes clear as soon as one takes into consideration that, in a way similar to English or Italian, the letter *gimmel* can be pronounced in Arabic as either /g/ or /J, zh/, depending on a dialect. In the Moroccan dialect of al-Idirisi it should be read as *Jajan* or even *Zhazhan*, which immediately reminds of Lake Zaysan lying along the course of the Irtysh river. That allows to identify the multiple Kimak settlements as being located on the shores of Lake Zaysan and along the Kara-Irtysh (presumably *Gamash* on the map, as if from a contracted pronunciation \**qa...ash*), where they were indeed supposed to be according the legend. This territory is designated on the map as *Ard-al-Kimakiyya* (The Land of the Kimaks). In reality, it most likely extended further to the northeast than the map shows, but Chinese Silk Road merchants rarely visited the northern tracks, so we see only its southern part.

Similarly, in the Muhamed al-Kashgari's sketchy drawing (c. 1072-74), we find the Yamaq Steppe positioned between the Ertish River and the Ili River (in the Tian Shan), therefore he also must have thought that the Kimak tribes lived somewhere between the Tian Shan and the Altai Mountains.

### Kimak phonology, grammar and lexis

Consequently, a matter that should be discussed in detail is the difference between the Kimak-Kypchak-Tatar, Kyrgyz-Kazakh, and Altay subtaxa, which are all frequently mixed up and intermingled in other classifications. How do these subtaxa differ? The following table shows that Proto-Kimak-Kypchak has undergone certain crucial transformations that made it phonologically very different from Kyrgyz-Kazakh and Altay, so they cannot be just blindly grouped together.

The Comparison of Differentiating Features in the Languages of the Great Steppe								
<b>Innovations in Proto-Kimak</b>	<b>Karachay</b>	<b>Typical Kimak- Kypchak-Tatar languages;</b>	<b>Baraba,</b> see [Dmitriyeva (1981)]	<b>Karakalpak</b>	<b>Kazakh</b>	<b>Kyrgyz</b>	<b>Standard Altay</b>	<b>English</b>

		see [Alishina (1992)], [Akhatov (1964)], [the Sibir Tatar lexicon was collected from a speaker on the net]						
	<b>KIMAK LANGUAGES</b>		<b>KYRGYZ LANGUAGES</b>			<b>ALTAY LANGUAGES</b>		
<i>Common Kypchak-Tatar innovative features not shared with Oghuz (blue, green)</i>								
The presence of the <b>intervocalic -w-</b> (either archaic or innovative)	Karachay <i>baur</i> <*bawir	Kazan Tatar <i>bawir</i> ; Bashkir <i>bawir</i> ; Sibir Tatar <i>pawir</i> ; Nogai <i>bavir</i> ; Kumyk <i>bavur</i> ;	Baraba <i>pawir</i>	<i>bawir</i> as in Kimak-Kypchak	<i>bawir</i> as in Kimak-Kypchak	<i>bo:r</i>	<i>bu:r</i>	liver
The presence of the <b>intervocalic -y-</b> (either archaic or innovative)	Karachay <i>süyek</i>	Kazan Tatar <i>söyäk</i> ; Bashkir <i>höyäk</i> ; Sibir Tatar <i>söyak</i> ; Nogai <i>süyek</i> ; Kumyk <i>süyek</i> ;	Baraba <i>süök</i>	<i>süyek</i> as in Kimak-Kypchak	<i>süyek</i> as in Kimak-Kypchak	<i>sö:k</i>	<i>sö:k</i>	bone
Differences in the <b>suffixes in "seed"</b>	Karachay <i>urluq</i>	Kazan Tatar <i>orliq</i> ; Bashkir <i>orloq</i> ;		<i>üriq</i>	<i>üriq</i>	<i>ürön</i> < Mong?;	<i>üren</i> < Mong?	see

		Sibir Tatar <i>orloq</i> ; Nogai <i>urlik</i> ; Kumyk <i>urluq</i> ;				cf. <i>uruq</i> "kin"	Also, in Khakas	
The use of <b>*bek</b> "very" in Kimak and <b>*ötö</b> in Kyrgyz-Kazakh	Karachay <i>bek</i>	Kazan Tatar <i>bik</i> ; Bashkir <i>bik</i> ; Nogai <i>bek</i> ; Kumyk <i>bek</i>	Baraba <i>bek</i> , <i>päk</i> ;	<i>zhüde</i>	<i>ötö</i>	<i>ötö</i>	<i>sürekey</i>	very (before adj)
<b>*oltur</b> versus <b>*otur</b>	Karachay <i>oltur-urGa</i>	Kazan Tatar <i>utir-irGa</i> ; Bashkir <i>oltur-urGa</i> ; Sibir Tatar <i>utir-ü</i> ; Nogai <i>oltir-</i> ; Kumyk <i>oltur-mak</i>	Baraba <i>oltir</i> , <i>otir</i> ;	<i>otiri-u</i>	<i>otir-u</i>	<i>otur-u</i>	—	to sit
<b>*ölön</b> versus <b>*ot</b> and <b>*chöp</b>	Karachay <i>hans</i> , <i>kirdiq</i>	Kazan Tatar <i>ölön</i> ; Bashkir <i>ülên</i> ; Sibir Tatar <i>ülên</i> ; Nogai <i>ölên</i> ; Kumyk <i>ot</i>	Baraba <i>öylän</i> , <i>ülän</i>	<i>shöp</i> , <i>ot</i>	<i>ot</i> , <i>shöp</i>	<i>chöp</i>	<i>ölön</i>	grass
<b>*qart</b> versus <b>*keri</b>	Karachay <i>qart</i>	Kazan Tatar <i>qart</i> ; Bashkir <i>qart</i> ; Sibir Tatar <i>qart</i> ;	Baraba <i>qart</i>	<i>Garri</i>	<i>kêri</i>	<i>qarî</i>	<i>qarGan</i>	old (person)

		Nogai <i>qart</i> ; Kumyk <i>qart</i>						
<i>*yilGa</i> versus <i>*özên</i>	Karachay <i>suu,</i> <i>qoban</i>	Kazan Tatar <i>yelga</i> ; Bashkir <i>yilGa</i> ; Sibir Tatar <i>yilGa</i> ; Nogay <i>yilGa suw</i> ; Kumyk <i>özen</i> ; <i>qoysuw</i>	Baraba <i>yilGa</i>	<i>özek</i>	<i>özen</i>	<i>özön</i>	<i>su:</i>	river
<i>*asha-</i> versus <i>*Je-</i>	Karachay <i>asha-rGa</i>	Kazan Tatar <i>ashau</i> ; Bashkir <i>ashau</i> ; Sibir Tatar <i>ashau, yeyü</i> ; Nogay <i>yew, ashaw</i> ; Kumyk <i>asha-maq</i>	Baraba <i>asha-</i>	<i>zheu,</i> <i>ishiu</i>	<i>zheu</i>	<i>zhesh</i>	<i>d'i:r</i>	to eat
<i>Common Kimak features also shared with Oghuz (blue)</i>								
An innovative <b>contraction</b> in "leaf" and "earth" (as in Oghuz)	–	Kazan Tatar <i>yafraq</i> ; Bashkir <i>yap rak</i> ; Sibir Tatar <i>yapraq</i> ; Nogai <i>yapirak</i> ; Kumyk <i>yapraq</i> ;	Baraba <i>yapraq</i>	<i>zhalbıraq</i> <i>topıraq</i>	<i>zhalbıraq</i> <i>topıraq</i>	<i>zhalbıraq</i> <i>topuraq</i>	<i>d'albıraq</i>	leaf earth



		Kazan Tatar <i>tufraq</i> ; Bashkirtupraq; Sibir Tatar <i>tuprak</i> ; Nogai <i>topıraq, topraq</i> ; Kumyk <i>topuraq</i> ;						
The innovative partial *S > y transition before open vowels (as in Oghuz)	Karachay <i>julduz</i> (/J/ as in Eng.) archaism	Kazan Tatar <i>yoldız</i> ; Bashkiryond ođ; Sibir Tatar <i>yoltos</i> ; Nogai <i>yuldız</i> ; Kumyk <i>yulduz</i> ;	Baraba *y-	<i>zhuldız</i>	<i>zhûldız</i>	<i>zhıldıs</i>	<i>d'ıldıs</i>	star
The -t/-d- : -l/-n- full softening in the verb suffix (as in Oghuz)	Karachay <i>jukla-</i> ; <i>ishle-</i> ;	Kazan Tatar <i>yoqla-</i> ; Bashkir <i>yoqla-</i> ; Sibir Tatar <i>yokla-</i> ; Nogai <i>uykla-</i> ; Kumyk <i>uykla-</i> ; Kazan Tatar <i>eshlêü</i> ; Bashkireshlêü; Nogai <i>êshlä</i> ; Kumyk <i>ishle-</i> ;	Baraba <i>yoqla-</i> (looks like a Kazan Tatar borrowing) Baraba <i>êshlä-</i>	<i>uyqıla-</i> <i>isleü</i>	<i>ûyıkta-</i> <i>istew</i>	<i>ukta-</i> <i>ishtö:</i>	<i>ishte:r</i>	sleep (v) work (v)
The -t/-d- :	Karachay		Baraba					

<p><b>-l/-n- softening</b> after consonants in the plural and accusative suffix (as in Oghuz)</p>	<p><i>-la, -lê;</i>  <i>-nu, -nü, -ni</i></p>	<p>Kazan Tatar <i>-lar, -lêr, -nar, -nêr</i> (plural); Sibir Tatar <i>-lar</i>; Nogai <i>-lar, -lêr</i> (plural)</p> <p>Kazan Tatar <i>-nî, -n</i>, (accusative); Nogai <i>-nî, -ni, -n, -dî, -di, -tî, -ti</i> ; Kumyk <i>-nî, -ni, -nu, -nü</i></p>	<p><i>-lar, -nar, -lär, -när;</i>  <i>-tar, -tär</i> (Radlov)</p> <p><i>-nî, -ni, -tî, -di, -ti;</i>  <i>-inî, -ini</i> (Radlov)</p>	<p><i>-lar, -ler,</i>  <i>-ni, -nî, -di, -dî, -ti, -tî</i></p>	<p><i>-lar, -ler,</i>  <i>-dar, der,</i>  <i>-tar, -ter,</i>  <i>-ni, -nî, -di, -dî, -ti, -tî</i></p>	<p><i>-lar, -ler, -lor, -lör,</i>  <i>-dar, der,</i>  <i>-dor, dör,</i>  <i>-tar, -ter,</i>  <i>-tor, -tör</i>  <i>-nu, -nü, -ni, -nî, -du, -dü, -dî, -dî,</i></p>	<p><i>-lar, -ler, -lor, -lör,</i>  <i>-dar, der,</i>  <i>-dor, dör,</i>  <i>-tar, -ter,</i>  <i>-tor, -tör</i>  <i>-ni, -nî, -di, -dî, -ti, -tî</i></p>	<p>the plural marker</p> <p>the accusative marker</p>
<p>The <b>-b/-p- : -m- softening</b> after consonants (as in Oghuz)</p>	<p>Karachay —  <i>kellik mise?</i></p>	<p>Kazan Tatar <i>ütmês</i>; Bashkir <i>ütmêβ</i>; Sibir Tatar <i>ütmês</i>; Nogai <i>ötpe</i>s; Kumyk <i>yaxshî ötmeygen</i>; Kazan Tatar <i>barasin mi?</i>; Sibir Tatar <i>paramisin?</i> Nogai <i>qördiN be?</i> Kumyk <i>geleJek mi?</i></p>	<p>Baraba <i>pu yiGit-mi?</i>  <i>kildi ba?</i>  (Radlov recorded <i>-b/-p-</i> in <i>-pîn, -bîn</i>"I am", which later mostly disappeared)</p>	<p><i>ötpês</i>  <i>keldi me?</i></p>	<p><i>ötpês</i>  <i>barasiN ba?</i></p>	<p><i>ötpöGön</i>  <i>keldi bi?</i></p>	<p><i>ötpös</i></p>	<p>dull (not cutting) question marker</p>

<p>The loss of <b>-Gaq</b> (as in Oghuz)</p>	<p>Karachay <i>qurGaq,</i> <i>quru;</i></p>	<p>Kazan Tatar <i>korï;</i> Bashkir <i>qoro;</i> Sibir Tatar <i>koro;</i> Nogai <i>kuri;</i> Kumyk <i>quru;</i></p>		<p>qûrGaq</p>	<p>qûrGaq</p>	<p>qurGaq</p>	<p>qurqaq</p>	<p>dry</p>
<p>The innovative voicing <b>t- &gt; d-</b> in some positions (as in Oghuz)</p>	<p>Karachay <i>tört</i> an archaism or back-mutation</p>	<p>Kazan Tatar <i>dürt;</i> Bashkir <i>dürt;</i> Sibir Tatar <i>türt;</i> Nogai <i>dört;</i> Kumyk <i>dört</i></p>	<p>Baraba <i>tört, dürt</i></p>	<p>tört</p>	<p>tört</p>	<p>tört</p>	<p>tört</p>	<p>four</p>
<p>The lack of the word-initial <b>m-</b></p>	<p>Karachay <i>burun</i></p>	<p>Kazan Tatar <i>borin;</i> Bashkir —; Sibir Tatar <i>poron;</i> Nogai <i>burin;</i> Kumyk <i>burun</i></p>	<p>Baraba <i>purin,</i> <i>murin</i></p>	<p>murin</p>	<p>mûrin</p>	<p>murun</p>	<p>—</p>	<p>nose</p>
<p><b>menen</b> versus <b>belen</b></p>	<p>Karachay <i>bla</i></p>	<p>Kazan Tatar <i>belen;</i> Bashkir <i>menê</i> <i>n;</i> Sibir <i>belen, men;</i> Nogai <i>-men;</i> Kumyk <i>bulan</i></p>	<p>Baraba <i>bilän, birlän, pilä, pirlän, pilan, pirlä, pila;</i> <i>minan, mina, ma:n;</i></p>	<p>menen, penen, benen</p>	<p>menen, -men, -pen; South Kazakh <i>pïpnan, -minan</i></p>	<p>menen</p>		<p>with someone</p>

The use of the <i>*achak</i> Future Tense	Karachay <i>-rik, -nik, -lik</i>	Kazan Tatar <i>-achak</i> ; Bashkir <i>-asaq</i> ; Nogai <i>-ayak, -eyek</i> ; Sibir Tatar —; Kumyk <i>-azhak, -ezhek</i> , Crimean Tatar <i>-aJak, -eJek</i>	Baraba <i>-är, -ir</i>	<i>-a-zhaq</i>	<i>-ar, -er, etc</i> <i>-maq, -mek,</i> <i>-baq, -bek</i> ( <i>-ayak, -eyek</i> only in western dialects)	<i>-ar, -er, etc</i>	<i>-ar, -er, -r;</i> <i>-at, -et</i>	Future Tense
The use of <i>*tegül</i> after adj. and nouns (as in Oghuz)	Karachay <i>tüyül</i>	Kazan Tatar <i>tügel</i> ; Bashkir <i>tügil</i> ; Sibir Tatar <i>tügel</i> ; Nogai <i>tuwil</i> ; Kumyk <i>tügül</i>	Baraba <i>tügül,</i> <i>tügil</i>	<i>emes</i>	<i>emes</i>	<i>emes</i>	<i>emes</i>	not
The absence of the word-final <i>-e</i> in <i>*tiz</i> ; and the use of <i>*tobuq</i>	Karachay <i>tobuq; tiz</i> (Balkar?)	Kazan Tatar <i>tez</i> ; Bashkir <i>tubiq</i> ; Sibir Tatar <i>tes, tubiq</i> ; Nogay <i>tiz</i> ; Kumyk <i>tiz(-ler), tobuq</i> ;	Baraba <i>tiz</i>	<i>dize</i>	<i>tize</i> ; cf. <i>tobiq</i> "ankle"	<i>tize</i>	<i>tize</i>	knee
The absence of <i>sizder</i> or <i>seler</i> (as in Oghuz)	Karachay <i>siz</i>	Kazan Tatar <i>siz</i> ; Sibir Tatar <i>ses</i> ; Nogay <i>siz</i> ; Cuman-Polovtsian <i>siz</i> ;	Baraba <i>sis, silär</i>	<i>siz</i>	<i>sender, sizder, siz</i>	<i>sizder, siler, sizler, siz</i>	<i>slerler</i>	you (plural)

		Kumyk <i>siz</i>						
The innovation <i>*nechik</i> versus <i>*qanday</i>	Karachay <i>qalay</i>	Kazan Tatar <i>nichek</i> ; Bashkir <i>nisek</i> ; Sibir Tatar <i>nitsek</i> ; Nogay <i>qalay</i> ; Kara Nogay <i>neshik</i> ; Kumyk <i>nechik</i>	Baraba <i>nê(n)chik</i>	<i>qalay</i>	<i>qalay</i> , <i>qaytip</i>	<i>qanday</i> , <i>qaytip</i>	<i>qandiy</i>	how?
The innovation <i>*quyash</i> versus <i>*kün</i>	Karachay <i>kün</i>	Kazan Tatar <i>qoyash</i> ; Bashkir <i>qoyash</i> ; Sibir Tatar <i>qoyash</i> ; Nogay <i>kün közi</i> ; Kumyk <i>gün(esh)</i> ,	Baraba <i>qoyash</i>	<i>kün, kuyas</i>	<i>kün</i>	<i>kün</i>	<i>kün</i>	sun
<i>*burada</i> < <i>*bu yerde</i> (as in Seljuk) along with the common and archaic <i>*munda</i> as in most TL's	Karachay <i>bilayda</i>	Kazan Tatar <i>biredê</i> ; Bashkir —; Sibir Tatar <i>piretê, pî vertê</i> ; Kumyk —;		<i>bul zherde</i>	<i>bûl zherde</i>	<i>bul zherde</i>	<i>bu d'erde</i>	here
The use of the verb <i>*is-</i> in reference to "wind" (as in Oghuz)	—	Kazan Tatar <i>isu</i> , Bashkir <i>iseu</i> ; Kumyk <i>esh-</i> ; <i>üfür-</i> ;	Baraba <i>ês-</i>	<i>zhiberou</i> , <i>yesiu</i>	<i>soGu</i>	<i>soGu:</i>	<i>soq</i>	to blow (wind)
<i>Other features</i>								

<p>The retention of the word final <b>-w</b> in <b>*suw</b>;</p>	<p>Karachay <i>suu</i></p>	<p>Kazan Tatar <i>sïw</i>; Bashkir <i>hïw</i>; Sibir Tatar <i>sow</i>, <i>sïw</i>; Nogaysïw; Kumyk <i>suw</i>;</p>	<p>Baraba <i>su</i></p>	<p><i>suw</i></p>	<p><i>su</i></p>	<p><i>su:</i></p>	<p><i>su:</i></p>	<p>water</p>
<p>The retention of the word final <b>-m</b> in "I'd rather do" versus <b>-n</b> in Kazakh-Kyrgyz;</p>		<p>Kazan Tatar <i>bara-yïm</i>; Bashkir <i>bara-inem</i> (?); Sibir Tatar <i>bara-yïn</i>; Nogay <i>bara-yïm</i>; Kumyk <i>bara-yïm</i>;</p>	<p>Baraba <i>bara-yïn</i>; <i>bara-yïm</i> (rare)</p>		<p><i>bara-yïn</i></p>	<p><i>bara-yïn</i></p>	<p><i>bara-yïn</i></p>	<p>I'd rather go</p>
<p><b>*ne(rse) de bulsa</b> in Kimak verus <b>*bir nerse</b> in Kazakh-Kyrgyz</p>		<p>Kazan Tatar <i>berär närsä</i>; <i>närsä dä bulsa</i>; <i>ni de bulsa</i>; Bashkir <i>berêy nêmê</i>, <i>nêmê bulha la</i>; Nogay <i>bir zat, ne di</i>; Kumyk <i>bir zat, ne busa da</i></p>	<p>Baraba <i>ällä nemä</i></p>	<p><i>bir närse, ne bolsa da</i></p>	<p><i>bir närse</i></p>	<p><i>bir nerse</i></p>	<p><i>ne de, neni de; neni-neni; bir neme</i></p>	<p>something</p>
<p><b>*kim-de</b> versus <b>*birö</b></p>		<p>Kazan Tatar <i>kem dä</i>; Bashkir <i>kem der</i>; Nogay <i>kim de</i>; Kumyk <i>kim busa da</i>; <i>bireu</i></p>		<p><i>bireu</i></p>	<p><i>bireu</i></p>	<p><i>birö</i></p>	<p><i>kem de</i></p>	<p>someone</p>

<p>The retention of the <b>word-final -sh</b>; with -s apparently being a local innovation that spread from Sibir Tatar and Nogai (?) into Kazakh</p>	<p>Karachay <i>tash</i></p>	<p>Kazan Tatar <i>tash</i>; Bashkir <i>tash</i>; Sibir Tatar <i>tos</i>; Nogay <i>tas</i>; Kумык <i>tash</i>;</p>	<p>Baraba <i>tash</i></p>	<p><i>tas</i></p>	<p><i>tas</i></p>	<p><i>tash</i></p>	<p><i>tash</i></p>	<p>stone</p>
---	---------------------------------	---	---------------------------	-------------------	-------------------	--------------------	--------------------	--------------

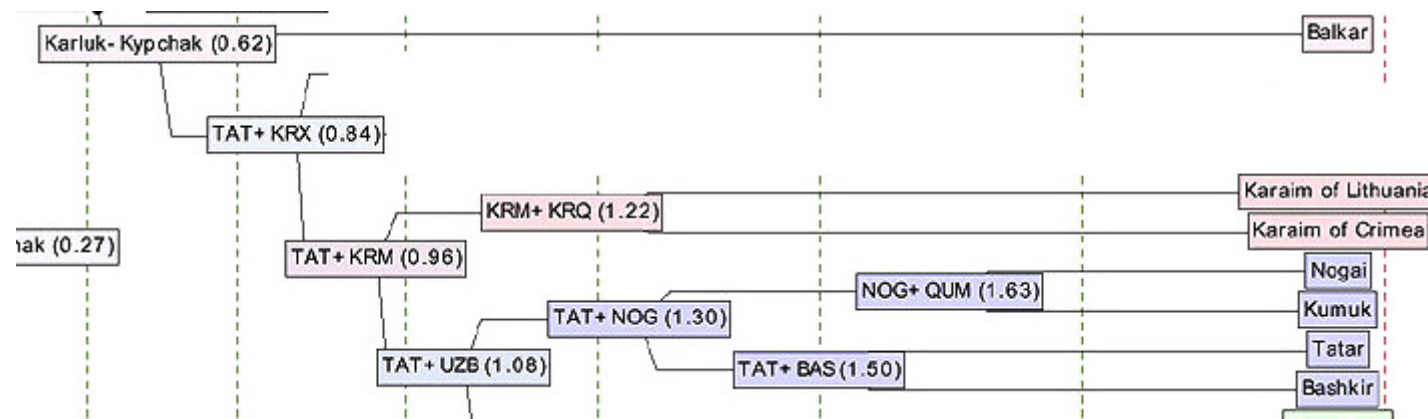
Evidently, this table demonstrates the differences between the Kimak-Kypchak-Tatar and Kyrgyz-Kazakh subtaxa, with Karakalpak being something of a secondary seam between the two of them.

#### Notes on other classifications and their positioning of Kimak

The table also shows why *Kazakh should be included into the same subtaxon with Kyrgyz*, whereas (Caspian) Nogai, on the contrary, has no direct bearing on either of them, and should be positioned into the same subtaxon as Kazan Tatar, unlike in an older Baskakov's classification. It is true, however, that Kazakh may exhibit some Kimak features, but these seem to stem from secondary contacts on the large territory of the Kazakh Steppe, which inevitably resulted in some intermingling of the early Kazakh speakers with the Kimaks.

Naturally, even more Kimak influence may be found in Karakalpak, which is essentially something of a northwestern variety of Kazakh.

Also, consider again the above-mentioned lexicostatistical research by Dybo (2006), which demonstrates the close proximity of some of the other Kimak-Kypchak-Tatar languages that were omitted in the present publication.



[Dybo, Anna, *The Chronology of Turkic Languages and the Linguistic Contacts of Early Turks* (2006)]

A similar classification had also been proposed at least as early as Bogoroditskiy (Kazan, 1934), unfortunately it was later superseded by that of Baskakov. Bogoroditskiy's classification was based purely on geographical principles, nevertheless it rather correctly differentiated (1) the many Khakas dialects; (2) the many Altai dialects; (3) the Siberian Tatars, e.g. Baraba; (4) Tatar, Bashkir; (5) Kazakh, Kyrgyz, Karakalpak, Uzbek, Uyghur; (6) Seljuk and Oghuz languages.

However, Baskakov (1960), apparently incorrectly, regrouped Kyrgyz with Altai, and Kazakh with Nogai, ignoring the obvious similarity between Kazakh and Kyrgyz, a view that lasted for about a half a century. Despite this and other similar drawbacks, Baskakov's classification was still the most detailed of its time.

For the above reasons, it is essentially incorrect to name both Kyrgyz-Kazak and Kimak-Kypchak-Tatar subtaxon as "Kypchak" (or "Kipchak" /keep-CHAHK ) as Baskakov and his followers tend to do. Initially, the term "Kypchak" seemed to refer only to a *relatively small clan* within the original Kimak confederacy. At a later stage, during the 11th-13th centuries this clan was present in many different parts of Eurasia, but that is just a different meaning of the term. The term "Kypchak" in the sense of tribal confederacy possibly referred to Cuman-Polovtsian or some of the Kimak tribes in contact with the Kievan Rus or just situated nearby, see for instance [*Gosudarstvo kimakov IX-XI vv. po arabskim istochnikam (The Kimak State of the 9-11th century according to the Arab sources)*, Kumekov, B.E.; Alma-Ata (1972)]. It actually takes a thorough historical study to explain who the Kipchaks were anyway, and Baskakov seems to omit this issue in his books.



Therefore we should assume that the term "Kipchak" originally had a much more narrow usage, until it was rather artificially attributed to all of the Great Steppe languages and more during the second half of the 20th century.

### **Conclusions:**

The **Kimak languages** *originally constituted a single linguistic unity that formed near Lake Zaysan and the upper Irtysh River by about 700 AD.*

*By c. 900 AD the Kimaks must have spread to the west across the Great Steppe territory and by 1050 AD reached the Kievan Rus.*

The term *Kimak* (sometimes named as "Kimak-Kypchak-Tatar" to keep some compatibility with the older terminology) may hereinafter be only applied to those languages which share the features described in the table above, and which therefore are particularly close to **Kazan Tartar**, the latter being a typical good example of modern Kimak languages. Other instances of Kimak languages include **Bashkir, Sibir Tatar, Mishar Tatar, (Caspian) Nogai, North Crimean Tartar, Lithuanian Karaim, Crimean Karaim, Kumyk**, possibly extinct **Cuman-Polovtsian**, and some other closely related dialects and languages.

The difficulties in the classification of Baraba (and particularly Tomsk) Tatars result from the scarcity of available materials, however **Baraba** seems to exhibit all the essential features of this Kimak subgroup just as well.

A special position belongs to **Karachay-Balkar** (see below).

These languages exhibit innovative features, which — as we shall explain in detail below — were mostly brought by their interaction with the Oghuz adstratum.

On the other hand, Kyrgyz, Kazakh and Karakalpak are more linguistically archaic and belong to a different subtaxon of the languages of the Great Steppe, named herein as the *Tian-Shan languages*.

One of the probable reasons why the Kimak languages finally grew so historically important may be connected to their *close original location to the northern track of the Silk Road* where they could interact culturally, linguistically and genetically with many different peoples and acquire certain knowledge and wealth that could have helped them to expand in the northwestern direction.

### **The relationship between Oghuz and Kimak**

### The Kimak and Oghuz secondary contact

Finally, we come to an interesting point mentioned above: *the Oghuz-Seljuk subtaxon seems to share some innovations with Kimak-Kypchak-Tatar*, namely:

(1) **the incomplete J- to y- mutation**, cf. Proto-Oghuz \**Jedi* "seven" attested by Mahmud al-Kashgari (see below), North Crimean Tatar *Jedi*, Kazan Tatar *Jide*, the intermingled allophonic use of *J/ y-* in East Bashkir dialects, etc., as opposed to the clear-cut Karakhanid *yeti*;

(2) **a sporadic t- to d- voicing**, cf. Gagauz, Turkish, Azeri, Turkmen *dört*, Kazan Tatar *dürt*, Nogai *dört* as opposed to the Karakhanid *tört*;

(3) **the loss of -G / -Gaq** as in Turkish *kuru*, Azeri *Guru*, Turkmen *Guri*, Kazan Tatar *korı*, Nogai *kuri*, as opposed to the Karakhanid *quruG* and Kazakh *qırGaq*;

(4) **a contraction in "leaf"** cf. Turkish *yaprak*, Azeri *yapraG*, Turkmen *yapraG*, Kazan Tatar *yafraq*, Nogai *yapırak*, as opposed to the Karakhanid *yapurGaq*;

(5) **the t : l transition** named herein as "*the heavy eastern versus the light western Turkic consonantism*", e.g. a "light" (lenitioned) *-l-* in the plural marker: *-lar* in Oghuz-Seljuk, Kimak-Kypchak-Tatar, Chagatai-Uzbek-Uyghur, Orkhon-Karakhanid, Khalaj, as opposed to the "heavy" (fortitioned) eastern pronunciation of *-dar/-tar-*, for instance in Kazakh-Kyrgyz, Baraba, Yugur and "Siberian" branches. Curiously, however, Kazan Tatar also preserves *-nar, -ner* which can be seen as an intermediate form between *-dar* and *-lar* as far as the degree of lenition is concerned. The stronger *-dar / -tar* and other fortified suffixes are also preserved in the East dialect of Bashkir (which was least affected by Kazan Tatar) as well as in Baraba. This may imply that the Kimak-Kypchak-Tatar languages originally had some phonological fortition typical of the eastern language clusters, whereas their historically recent lenition is probably acquired from Oghuz;

(6) **the use of \*tegül instead of e(r)mes**, cf. Turkish *deGil*, Azeri *deyil*, Turkmen *del*, Kazan Tatar *tügel*, Kumyk *tügül* as opposed to the Karakhanid *ermes*, Kazakh-Kyrgyz *emes*;

(7) **the use of the \*aJak in Future Tense**, cf. Turkish *-aJak/-eJek-*, Turkmen *-Jak/-Jek*, Kazan Tatar *-achak-*, Bashkir *-asaq-*, Nogai *-ayak/-eyek-*, Crimean Tatar *-aJaq/-eJeq-*, Kumyk *-azhaq/-ezhek*. The tense is also used in Karakalpak in the Aral-Caspian region probably because of the Oghuz (Turkmen) presence there;

(8) **the frequent use of -dir/-tİR in the 3rd person singular**, cf. Turkmen, Azeri, Turkish; Cuman-Polovtsian, Kazan Tatar *-dir/-tİR*, etc. as opposed to its absence in Kazakh and Kyrgyz at least as far the copula construction is concerned (e.g. *Ol qazaq* "He is a Kazakh), etc;

On the other hand, despite this presumable relatedness, presently there is only poor mutual intelligibility between modern Oghuz-Seljuk and Kimak-Kypchak-Tatar languages, with many differences in syntax, morphology and semantics. With the 70% of average similarity between Turkmen and the modern languages of the Golden Horde, the present-day distance between even the most archaic and easternmost Oghuz languages and the Kimak-Kypchak-Tatar languages seems to be rather considerable.

For instance, with the 65% between Turkish and Tatar in Swadesh-215 (borrowings excluded), the actual difference in real speech would normally be considerably beyond comprehension. A few simple phrases from Tatar-Turkish phrasebook may look as follows:

Kazan Tatar *Sin kay-a bar-a-sin cong?* cf. Turkish *Sen nere-ye gid-i-yor-sun?*, literally "You where going-are-you?";

Kazan Tatar *Salkin su bir-egez-che* cf. Turkish *Souk su ver-in (lütfen)*, "Cold water give-please";

Kazan Tatar *Gailê-biz-de öch bala — min, apa-m hêm ene-m*, cf. Turkish *Aile-miz-de üç chojuk (var) — ben, abla-m ve (hem de) kardesh-im*, "Family-my three child — me, sister-my and brother-my".

That does not mean, of course, that Kimak and Oghuz have nothing in common with each other, it is just that the described changes seem to be roughly consistent with at least 1500-2000 years of glottochronological separation, which makes the recent existence of an Oghuz-Kimak genetic unity an unlikely option.

And indeed, as we will conclude below, the phonology, grammar and particularly the vocabulary of Oghuz languages are in good correspondence with Karakhanid, taken that that Proto-Oghuz originally belonged to the same stock as Orkhon Old Turkic, Old Uyghur and Karakhanid, which seems to refute the above idea of Oghuz-Kimak relationship.

But if Oghuz and Kimak are not really close, where do these shared elements come from, anyway?

We may not suppose that these could have emerged independently in each subtaxon, since the coincidence of several simultaneous mutations is statistically negligible, therefore a much more likely and interesting option would be that they occurred due to the *secondary contact and mutual intermingling, when at some point in time, the early Oghuz tribes crossed the area of the Kimak tribes.*

### **The hypothesis of linguistic exchange in northern Kazakhstan**

The conclusion of secondary relatedness between Kimak and Oghuz is in accordance with the historical records saying that Seljuk's clan separated from the Transoxanian (=Aral-Caspian ) Oghuz tribes near the Syr-Darya in the Kazakhstan steppe, which seems to have been the traditional habitat of the Kimak-Kypchak-Tatar or Kazakh tribes. In other words, it is geographically simple to assume that the Oghuz and the Kimaks, being so geographically close, might have formed a sort of a linguistic area near the Aral Sea. Curiously, Al-Kashgari claims that "*Kirkiz, Kifzhak, Uguz, Tuxsi, Yagma, Jikil* [the latter three tribes apparently were located near the Ili river in the Tian Shan], *Ugrak, Jaruk all have one pure Turkic language. Close to them are the dialects of Yamak [= probably Kimak] and Bashkirt...*", which evidently positions "*Uguz*" into the same geographic and linguistic row as Kyrgyz and Kypchak with several lesser medieval tribes.

We can also find multiple historical records mentioning a Kimak-Oghuz alliance in the 10th century. For instance, Arab geographer Al-Masudi wrote c. 930 that the Kimaks and Oghuzes were *coaching* along the Emba and Yaik together.

**Note:** the English word *coach* is from French, where it seems to go back to Hungarian, where it is probably from Bulgaro-Turkic *\*köch-* "to migrate" [Webster's New World Dictionary (1986), Sevortyan's Dictionary (1980)]

Ibn Haukal c. 950 drew a map showing that Kipchak-Kimak tribes together with the Oghuz tribes were pasturing their cattle in the steppes north of the Aral Sea. Al-Biruni c. 1000 noted that Oghuz tribes quite often pastured in the country of the Kimaks [en.wikipedia.org].

However this hypothesis does not explain why the above-listed features passed into nearly all of the Kimak languages, which implies that the actual interaction must have occurred much earlier when both Kimak and Oghuz tribes were still living in the same relatively small area, such as a passage between mountain ranges, so their linguistic contacts must have been very intense and taking place at the proto-language level. For this reason, below we will consider another hypothesis that suggests a cultural and linguistic exchange near Lake Zaysan.

### **The hypothesis of linguistic interaction near Zaisan**

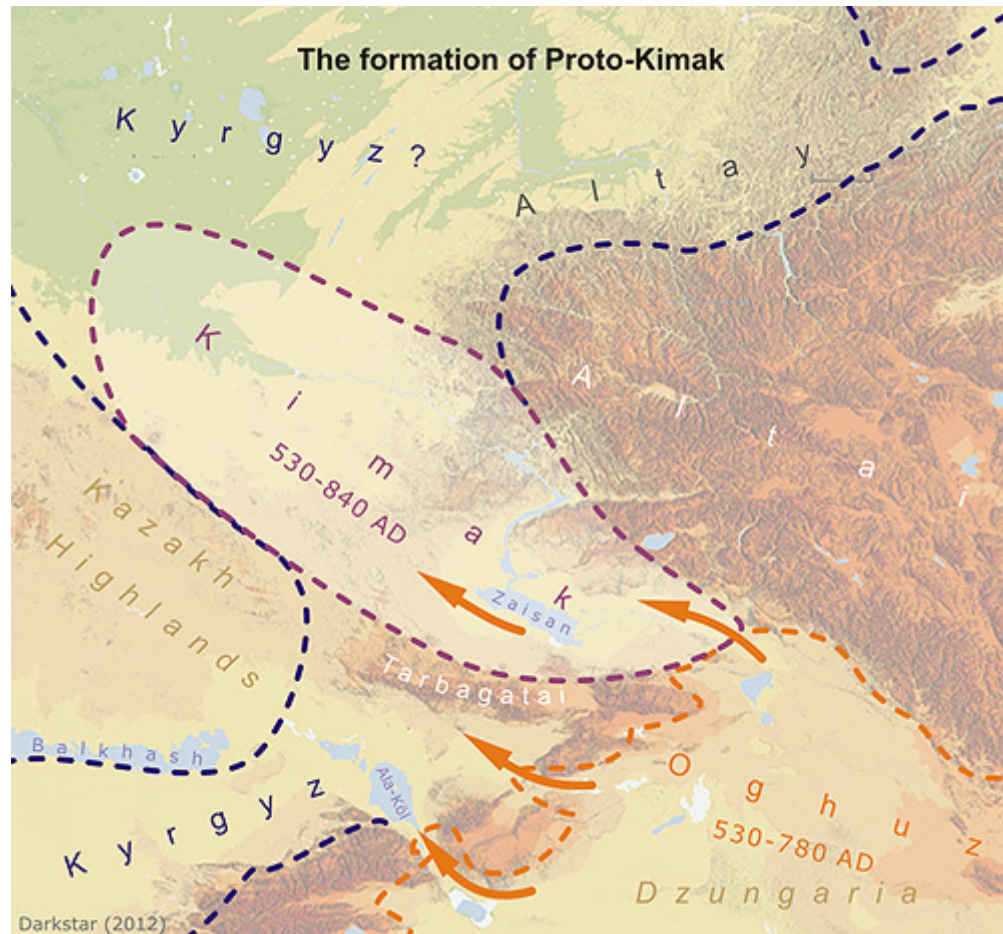
Beginning of 552 AD some of the Great-Steppe tribes were subdued by the western Göktürks, who essentially must be the speakers of an unidentified Orkhon-Oghuz-Karakhanid dialect, such as Old Uyghur or Oghuz judging from their geographic position near Dzungaria. Presumably, this West Göktürk language-dialect must have acquired a high sociolinguistic status in many Turkic-speaking societies of the time.

It is quite plausible to assume that Proto-Oghuz could have actually formed a considerable part of that West Göktürk dialect area given its later tendency to migrate in the western direction along the same path.

Initially, Proto-Kyrgyz was a conservative Turkic language apparently distributed either (1) along the Irtysh or (2) between the Irtysh and Ob rivers, essentially in the area known as the Baraba and Kulunda Steppe, or (3) in the area between the Altai and Tian Shan Mountains.

Whereas Proto-Kyrgyz-Kazakh had occupied the area west of the Altai Mountains and east of the Tian Shan for many centuries, Proto-Oghuz was probably a recent arrival from Dzungaria brought by the expansion of western Gökturks after 530-550 AD.

Consequently, we can infer that *somewhere around 550-800 AD there occurred a strong linguistic exchange between Proto-Oghuz in Dzungaria and the early Kyrgyz dialects north of the Tarbagatai in the Great Steppe, which could have resulted in the formation of Proto-Kimak.* In other words, the most simple and plausible hypothesis which would explain all the relations among Proto-Oghuz, Proto-Kimak, and Proto-Kyrgyz-Kazakh, would be that *the area of Proto-Kimak must have originally formed as a transitional region where the early Kyrgyz dialect overlapped and intermingled with Proto-Oghuz.*



The map of Proto-Oghuz and Proto-Kyrgyz hypothetical exchange between 550-800 CE

The overlapping of the Oghuz Kyrgyz area soon resulted in the formation of a new transitional dialectal seam, which became known as Kimak. This Kimak area shared archaic linguistic features with Kyrgyz, on one hand, and some innovative features with the early Oghuz, on the other.

Furthermore, *Oghuz too was affected by Kimak and Kyrgyz dialect-languages*; it absorbed some of their elements, to some extent even becoming part of the Great Steppe Sprachbund, and deviating from its Orkhon-Karakhanid parent stem.

On the other hand, *the speakers of Kyrgyz were largely unaffected by the Göktürk dialect-languages* because these were already absorbed and buffered in the Kimak zone. Consequently, the Proto-Kyrgyz-Kazakh-Uzbek-Uyghur language became locked in a sort of linguistic refugium near the foothills of the Tian Shan Mountains where it was able to retain many of the archaic features from before the 6th century.

### **Conclusions:**

As the Western Göktürk tribes, apparently speaking a language similar to the early Old Uyghur, moved back from Mongolia into the upper reaches of the Irtysh river between 550-700 AD, they must have come into contact with the local Proto-Kyrgyz tribes. This intermingling must have resulted in the formation of the three local dialectal areas:

(1) **Proto-Kyrgyz (or Proto-Tian-Shan)** (possibly also including Proto-Karluk): this area that was almost unaffected by the Göktürk language ultimately led to the emergence of the now-extinct Karluk (uncertain), the Tian-Shan Kyrgyz, and finally, after the 15th century, Kazakh and Karakalpak languages;

(2) **Proto-Kimak:** this area was strongly affected by the Oghuz or Western Göktürk migration, but retained many older Kyrgyz elements, for instance *-w-* in *bawir* "liver", and *-w* in *taw* "mountain", as opposed to the *-G-* and *-G* in the oncoming West Göktürk language — to name just a few typical features;

(3) **Proto-Oghuz:** this area acquired certain features from Kimak, but otherwise remained relatively unaffected, retaining many Orkhon-Karakhanid archaisms from an older period.

### **On the origins and history of the ethnonym *Tatar***

Speaking of the earliest clear-cut attestation of the ethnonym *Tatar*, we should probably turn to the Orkhon Turkic inscription of Kul Tegin made in 732, which cites a reference to the burial of Bumin Kagan in 552. The attestation consisted of the following passage, "...*Böküli Chölüg* (=the Koreans), *TabGach* (=the Chinese), *Avar*, *Rome* (=the Byzantines), *Kirgiz*, *Uc-Quriqan* (=the Proto-Yakuts), ***Otuz-Tatar***, *QitaN* (= the

*Khidans = the Mongolic peoples in the Greater Khingan Mountains) and Tatabi, this many people came...*" [see [Türük Bitig](#), a site dedicated to Orkhon-Yenisei inscriptions].

This suggests that *by 550 AD the Tatars constituted a political or military confederacy made up 30 (otuz) different clans or tribes and probably united as one single kaganate, though their exact location is unknown.*

**Note:** Herein we are trying to consistently exclude any early evidence from Middle Chinese records due to their ambiguity and multiple difficulties with the verification and interpretation. However, according to the Chinese version, the word *ta-da* or a similar one could have been initially used as the Chinese exonym applied to all of the foreign tribes beyond the Great Wall, similar to the *barbars* of the Greeks.

Moreover, and quite confusingly, the Tatars are described in the *Secret History of the Mongols* circa the 1190's, living somewhere near the modern-day border of Buryatia and Mongolia along the Onon River (which is the tributary of the Amur, and being the sworn enemies of Genghis Khan). Those Mongolian Tatars had poisoned his father and waged war on Genghis Khan, but then were finally exterminated in retaliation when he came to power.

The *History* does not explain which language they spoke, whether they were Turkic or Mongolic, it only suggests that they were able to say at least a couple of phrases in Middle Mongolian. More curiously, the two names of Genghis Khan himself, the original one *Temüjin* created after the name of a Tatar *Temüjin-üge* — presumably from Turkic *Temir-ji Aga*"The Blacksmith Brother — , and the later one *Jenghis Kagan*, probably chosen after a certain *Lake Tenghis* mentioned in the first lines of the *History* (Turkic "The Sea", probably Lake Baikal), both indicate the existence of Turkic ethnonyms and toponyms in the area, which may finally mean that these Mongolian Tatars, vividly described by Genghis Khan and his court scribes, were indeed of Turkic origin [see *the Secret History of the Mongols* (1240), translation by F. W. Cleaves from the Mongolian original (1982)].

Judging from their location in the Trans-Baikalian region, we may suppose that these Tatars could in fact have been a lost extension of Proto-Sakha, most likely related to Kurykans, who had integrated into the local Mongolic society (and possibly adopted the Mongolian language).

According to the legend cited by Gardezi (1030) and described in the chapter about *The Kimak subtaxon*, the ethnonym *Tatar* is also clearly traceable to a certain clan within the Kimak Confederacy situated along the Irtysh River circa 700 AD.

Consequently, one may wonder about at least three different early mentions of Tatars in three different contexts — one before the formation of the Kimak confederacy, another one as a part of it, and yet another one in reference to the purported Turkic tribes of Mongolia and Trans-Baikalia. What is the difference among the three?



As explained in the chapter about the Turkic ethnonymy, the most likely hypothesis about the Tatar origins would be that *the word Tatar must have originally been the name of a patrilineal clan working as a sort of equivalent of a European surname*. In other words, this hypothesis suggests that the word *Tatar* may originate in a personal name or alias of the Tatar clan's progenitor. (But what this name or alias could have initially meant would be just anybody's guess.)

Consequently, when the legend teller says that the men named *Tatar, Kimak, Kipchak*, etc. came over to live with the man named Shan, he probably just means that these could either be their original first names in some cases or their preexisting clan surnames in others.

Since the patrilineal clan of Tatars and the surname of Tatar may have merely genetic but not necessarily linguistic connection to its members, any men who belonged to that clan could have possibly spoken a generally unknown Turkic dialect or even a Mongolic language and lived in unspecified parts of Eurasia.

We cannot even exclude the possibility that some of the Tatars may have deliberately adopted their surname under generally unknown circumstances, even though they were not genetically connected to the original clan of Tatars. The existence of Mongolian Tatars described in the *Secret History of Mongols* is particularly interesting and questionable in this respect.

However, we should assume that most European and West Siberian Tatars, that the ethnologists are usually familiar with, supposedly trace their patrilineal descent (1) either to the Tatar man of the Kimak Confederacy, who had no first name and who settled down with Shan of the Tatars circa 700 AD, or (2) to Shan himself, or (3) they both were the same person, the latter option being the most simple and plausible one.

If the Mongolian Tatars indeed were of Proto-Sakha origin, then their separation from other Tatar clans could have occurred at the Proto-Turkic level, somewhere before 1000 BCE because of the very early separation of Sakha, which would make Tatars one of the earliest attested Turkic clan.

As for the rest of it, the actual use of this word *Tatar* throughout history has been quite different and variable — rising from the limited, regional usage as a clan name to an all-encompassing Turkic and Mongolic exonym and then falling into disuse again.

In 922, the "al-Bashkird" of Ibn-Fadlan were already attested near their present-day location west and southwest of the Urals, however there is no direct reference to the Kimak-related Tatars, as yet. Presumably, in the course of the 9th-10th centuries, during the period of the Kimak dissemination over the Great Steppe, the Kimak Tatars must have become the ruling clan among the Kimaks.

As one may suppose, during that period the word *Tatar* must have gained a socially prestigious connotation of a leading clan's title, and many Kimaks might have attempted to trace their personal roots specifically to Tatars. That honorific usage could have lasted well into the times of the

Mongols in the 13th century, so finally the Mongols themselves were frequently conflated with the Tatars. Giovanni da Pian del Carpine (1245), for instance, consistently names all the Mongols as Tatars despite his personal visit to Mongolia.

This ethnonymic confusion can also be explained from the military standpoint: the aristocracy of Mongolic descent constituted only a small part of the Golden Horde population, at least during its later stages, and the Mongolic tribes had initially been far too small to achieve the conquest of the enormous territory they acquired. Therefore, it is implausible that the Mongol generals were able to do without any help from the locals, they must have recruited the regional Turkic population into their armies, most of whom were evidently of Kimak-Kypchak-Tatar origin. Therefore, the actual conquest and control over the land was probably achieved by means of the ruling Tatar clans. However, there are few specific historical documents that could corroborate this outlook.

According to a different version [sources and details?], the name Tatar was brought only during the Mongolian period.

The ethnonym *Tatar* was particularly widespread among the Golden Horde aristocracy, military and local officials [see for instance *The Great Russian Encyclopedia* (2004)]. The linguistic differentiation among the Turkic dialects of the Golden Horde was evidently small, so all of the Golden Horde peoples between the 13-17th centuries were collectively called *Tatars* in Russia, many parts of Central Asia and Europe.

In Latin-speaking Europe, the word Tatar was frequently changed to "*Tartar*", apparently due to the association with the *Tartarus*, which, according to Greek mythology, was the underworld at the bottom of the abyss beneath the earth, where an anvil takes nine days to fall.

After the dissolution of the Golden Horde, the term must have acquired negative connotations, whereas many post-Golden-Horde ethnicities came up with other newly-formed names, such as *Noghai* (=from the Noghai Khanate, after the name of a Mongol general), *Mishar*, *Kazanly* (=from the Kazan Khanate), etc. For instance, in reference to the 18th-19th century, Carl Ritter, citing the research of German ethnographer Julius Klaproth (1783–1835), notes the following:

*"But if you ask the so called Kazan or Astrakhan Tatar, if he is a Tatar, he will answer negatively, for he names his dialect 'Turki' or 'Turuck', not 'Tatar'. Being aware that his ancestors were subdued by the Tatars and Mongols, he takes the word 'Tatar' as pejorative and meaning nearly the same thing as a bandit." [See *Die Erdkunde im Verhältniss zur Natur und zur Geschichte des Menschen (Geography in Relation to Nature and the History of Mankind)*, written 1816–1859]*

During the period of Ivan the Terrible (1530-84), who moved the imperial frontier beyond the Ural Mountains, the ethnonym *Tatar* was presumably carried further into Siberia by Russian Cossacks. Supposedly, this is how it came to be applied to the Sibir Tatars of the Tobol-Irtysh area, the Baraba Tatars, the Altay Turkic peoples and the Yenisei Kyrgyz tribes of the 17th century, though the presumable Russian origin of the

Tatar self-reference among these people is disputable. In any case, until the beginning of the 20th century, the Altay-Sayan peoples were known under such names as *Abakan Tatars*, *Chulym Tatars*, *Kuznetsk Tatars*, *Azerbaijani Tatars* and so forth.

Only the Kyrgyz and the Ottoman Turks were among the few that never received this exonym.

By the 18th century, the name became so overextended and overused, that it began to include any people of East Asia. French Sinologist Abel-Rémusat, for instance, used the term "*Tartares*" as a catch-all name for "*des Mandchous, des Mongols, des Ouigours et des Tibétains*" as late as 1820.

Moreover, until the 19th century, Siberia was often designated as *Tartaria (Magna)* in Latin or *Grande Tartarie* in French or *Tartary* in English on most geographic maps, see, for instance, Nicolaes Witsen, *Noord en Oost Tartarye...*, (1672). In other words, the expression *Tartaria (Magna)* was used in the same way as *Siberia* today.

Hence, also the name of the Strait of Tartary between mainland Russia and Sakhalin Island. The name was coined by La Perouse in 1787, even though no Turkic peoples had lived there ever.

During the reign of Peter the Great (1682-1725), when Turkology began to rise as a distinct branch of science in the Russian Empire and Western Europe [see Baskakov, *N.A. Vvedeniye v izucheniye tyurkskikh yazykov (An introduction into the study of Turkic languages)*, (1969); chapter *The history of study of Turkic languages in Russia before the 19th century*, p. 18], nearly all the known Turkic languages and dialects (outside Ottoman Turkish) became generally known as *tatarskiye narechiya* "Tatar dialects" in Russian. And, in some cases that indiscriminately included Mongolic, Tungusic, Tibetic, Samoyedic and other completely unrelated Siberian ethnic groups.

Strahlenberg and Messerschmidt (1720-1730), the earliest European explorers of Siberian peoples, were apparently a little unsure about the proper usage, however Strahlenberg [*Das Nord und Ostliche Theil von Europa und Asien*, Stockholm, 1730 ] seems to use the word *Tataren* as a generic term for the Turkic-speaking peoples only, not Mongols or anyone else.

The *Brockhaus and Efron Encyclopedic Dictionary* (1906), widely popular before and even after the Russian Revolution, openly protested against that overused terminology,

*"Tatars do not exist as a single ethnicity; the word "Tatar" is nothing but a collective nickname for a number of peoples of [sometimes] Mongolic, but particularly Turkic descent, speaking Turkic languages, and of Quranic affiliation. [...] From scientific perspective, the name of Tatar has presently been rejected when applied to Mongols or Tunguses, and retained only in reference to those linguistically Turkic ethnicities that form part of the Russian Empire, but excluding other Turkic nations with independent historical appellations (Kirgizes, Turkmens, Sarts, Uzbeks, Yakuts, etc). Certain scientists (Yadrintsev,*

*Kharuzin, Shantr) have suggested to modify the appellation terminology of some of the Turco-Tatar ethnicities [...], for instance, by renaming Azerbaijani Tatars to Azerbaijanis, Altay Tatars to Altayans, etc., but that has not gained much acceptance, as yet [...]"*

As a result, the indiscriminate term *tatarskiye narechiya* "Tatar dialects", generally accepted in the 19th century, was soon supplanted by the names of specific languages that appeared during the 1920-30's post-revolutionary renovation, though in some cases, such names as Uzbek, Uyghur, Khakas seem to have been taken right off the top of the head and then granted by consensus.

For some time after the revolution, "*Turkish-Tatar languages*", "*Turkish languages*", "*Turco-Tatars*" were still variably used as generic terms by various authors between the 1800-1930's . But aAfter the rise of the Republic of Turkey (1922) and its frequent generalization of *Türk* as a comprehensive, far-reaching concept, the recognition of the newly-formed term *tyurkskije jazyki* "Turkic languages" must have finally become widespread and generally-accepted even in reference to the ethnic groups that never called themselves *Turks*.

Nevertheless, the older usage in such phrases as *tataro-mongoly* "Tatar-Mongols" or *tataro-mongolskoye igo* "Tatar-Mongol yoke", referring to the rise of the Golden Horde and its punitive raids against Rus, still exists in Russian historiography.

Apparently, the extensive use of the term *Kypchak* popularized by Baskakov's classification (1950-1980's) followed the same avoidance strategy by trying to get rid of the word *Tatar*. As a result, in certain contexts, both names became nearly synonymous, the former being sort of euphemistic for the latter.

In the beginning of the 21st century, the name Tatar is formally retained mostly just by the **Kazan Tatars of Tatarstan** (who sometimes object to its usage), **Crimean Tatars**, **Mishar Tatars** west of Tatarstan, **Sibir (Tobol-Irtysh) Tatars** (whose language is poorly documented in the scientific literature), **Baraba Tatars** (on the verge of linguistic extinction, but often just "Baraba"). It is also accepted as a generic self-appellation *Tadarlar* by various Khakas and Altay Turkic ethnicities, and sometimes can be applied to other smaller and lesser-known ethnic groups, such as Astrakhan Tatars, Lithuanian Tatars, etc.

### **Bashkir is closely related to Kazan Tatar**

Judging solely by a superficial look at the orthographic phonology, a casual onlooker may think that Bashkir might be a strongly differentiated language among Turkic, no less than Chuvash or Sakha. However, at closer examination, *one can find a remarkable lexical similarity of more than 95% between Kazan Tatar and Bashkir in Swadesh-215*. A significant error in this figure is rather unlikely, taken that the list was composed by proficient speakers at Wiktionary.org and then rechecked through dictionary search herein.

The few clear-cut lexical and semantic discrepancies found in Swadesh-200 are as follows:

<b>Bashkir</b>	<b>Kazan Tatar</b>
<i>tubiġ</i> "knee"	<i>tiz</i> "knee"; <i>tubik</i> "ankle";
<i>tanau</i> "nose"	<i>borin</i> "nose"; <i>tanau</i> "muzzle"
<i>ésê(y)</i> "mother"	<i>ana</i>
<i>nimê</i> "what"	<i>nêrse</i>
<i>saN</i> ( tuZan	

"dust"*tuzan aliS* (alus , als)), *yïraq* "far"*yeraq* usually *bisraq* "dirty"*shaqshî, kerle, pïchraq bïnda* "here"*mïnda, biredê* "here", with the latter word obviously from Oghuz, cf. Azeri, Turkish *burada*

Despite the similarity, there may be more lexical differences that are less distinct, such as different semantic connotations of the same word, synonyms, slightly different phonology, etc.

Moreover, the speakers of both languages report good mutual intelligibility, even though the Bashkir phonology developed some remarkable innovations which in any way can hardly be any more pronounced than, say, those in northern British and American English.

As a result, the terms *Bashkir Tatar* and *Kazan Tatar* would be more self-explanatory for educational purposes, though the general trend is to drop the "*Tatar*" ending, not to add it.

Curiously, unlike the English dialects, the odd phonology of Bashkir is hardly notable in real speech, and practically speaking, Bashkir has almost the same "sound" to a casual listener as Tatar, Kazakh and other languages of the Great Steppe. This is an interesting example how misleading the observations of orthographically-reflected written phonology from textbooks alone can be. Compare a similar situation with Uzbek-Uyghur where phonology points in the direction of Karakhanid, while everything else points to Proto-Kyrgyz-Kazakh.

As far as the phonological laws are concerned, note the typical innovative vowel mutations in Kazan Tatar and Bashkir that often set them aside from the nearby Kimak languages:

- (1) the *i* > *e* **vowel mutation**, as in Kazan Tatar and Bashkir *tel* "tongue"; *bel-* "know"; *ber* "one";
- (2) the correspondent circular *e* > *i* **vowel mutation**, as in Kazan Tatar and Bashkir *it* "meat"; *ni* "what?";
- (3) the *u* > *o* **vowel mutation**, as in Kazan Tatar *ozin*, Bashkir *ozon* "long"; Kazan Tatar *bolit*, Bashkir *bolot* "cloud";
- (4) the correspondent circular *o* > *u* **vowel mutation**, e.g. in Kazan Tatar and Bashkir *urman* "forest"; *qul* "arm, hand"; *ut* "fire", etc.

These vocalic mutations are rather unique among the Turkic languages. The fact that they are noticeable mostly in vocalism is indicative of the recent separation of two languages, since vowels tend to change faster than consonants.

### **On the origins of the ethnonym "Bashkort"**

The autonym *Bashkort* is often explained as Turkic *bash* "head" + Oghuz *kurt* "wolf", where *kurt* is euphemistic for "wolf" though originally meaning "worm, bug". However, in modern Bashkir, *qort* in fact means "larva", so the immediate meaning poises questions concerning the origins of the ethnonym.

The word *kurt* with the meaning "wolf" is actually a purely Oghuz word, evidently with the original implication "a parasite that kills the sheep"; it is also sometimes thought to be influenced by Persian and West Iranian *gorg* "wolf". The use of an Oghuz word instead of the original Bashkir word *büre* (common to many Turkic languages) too raises doubts about the correctness of this interpretation.

We know that the Bashkort people were mentioned in several Arab sources since c. 840; at that time, they were said to occupy the territory south of the Ural Mountains — from the Volga and Kama to the Tobol River. Ibn-Fadlan clearly mentions certain "al-Bashkird" located in the present-day Tatarstan near the Kama River as early as 922, he says, "*We arrived in the land of the Turks called al-Bashgird... these were the most foul of all the Turkic peoples... when one of them meets a man, he cuts his head...*". He also found them near the Emba River (to the south of the Urals), which is evident from his words, "*...to protect them [=the carts] from the Bashkir(d)s in case they capture them...*".

Hence, we can infer that the name originally referred to a "headcutter (-splitter, -buster)" > caravan robber, and could have been *ambiguously applied to various robbers and cutthroats* from Kimak-Kypchak-Tatar groups distributed around the Urals, but was unluckily retained into the modern period only by the modern Ural Tatars (Bashkirs). Again, the practice of killing strangers was widespread in many early societies, it is mentioned for instance for the neighboring Mordvins of the 13th century [see the writings of Friar Julian (1235) below].

The name could also have referred, just as in many other Turkic clans, to the name or alias of the hypothetical clan's progenitor. Originally meant to imply force and fury, and the ability to defend against the enemies, such an implication must long have become unacceptable, and its primary meaning must have been forgotten.

Moreover, one can easily note that there is certain geographical discrepancy of about a hundred miles in the location of Ibn-Fadlan's *al-Bashkird* (which were mentioned in two areas: the present-day Tatarstan and the area along the Yaik river) and the modern Bashkortostan (which is situated in the Southern Ural). This indicates that Ibn-Fadlan, as well as other Arab historians and travelers, apparently used this ethnonym to refer to what we would presently call "Proto-Kazan-Tatars", "unidentified Kypchak tribes" or at least "the southern and western Proto-Bashkirs". This suggests that at least before the 13th century, *Bashkird* was in fact a popular early ethnonym for many different Tatar-Kipchak groups scattered from the Volga to the Ural mountains, but was retained into present only in the Ural Mountains, which served as a sort of the ethnonymic refugium for this name.

### **The Proto-Hungarian influence in Bashkir**

The habitat of the present-day Bashkir people matches the area of a South Ugric substratum (the extinct South Mansi languages) and probably even the territory of Magna Hungaria, the supposed Proto-Hungarian Urheimat.

The people in that area were still mentioned to speak a sort of Proto-Hungarian as late as 1235 shortly before the arrival of Mongols. [Friar Julian](#) is said to have discovered the following in this respect:

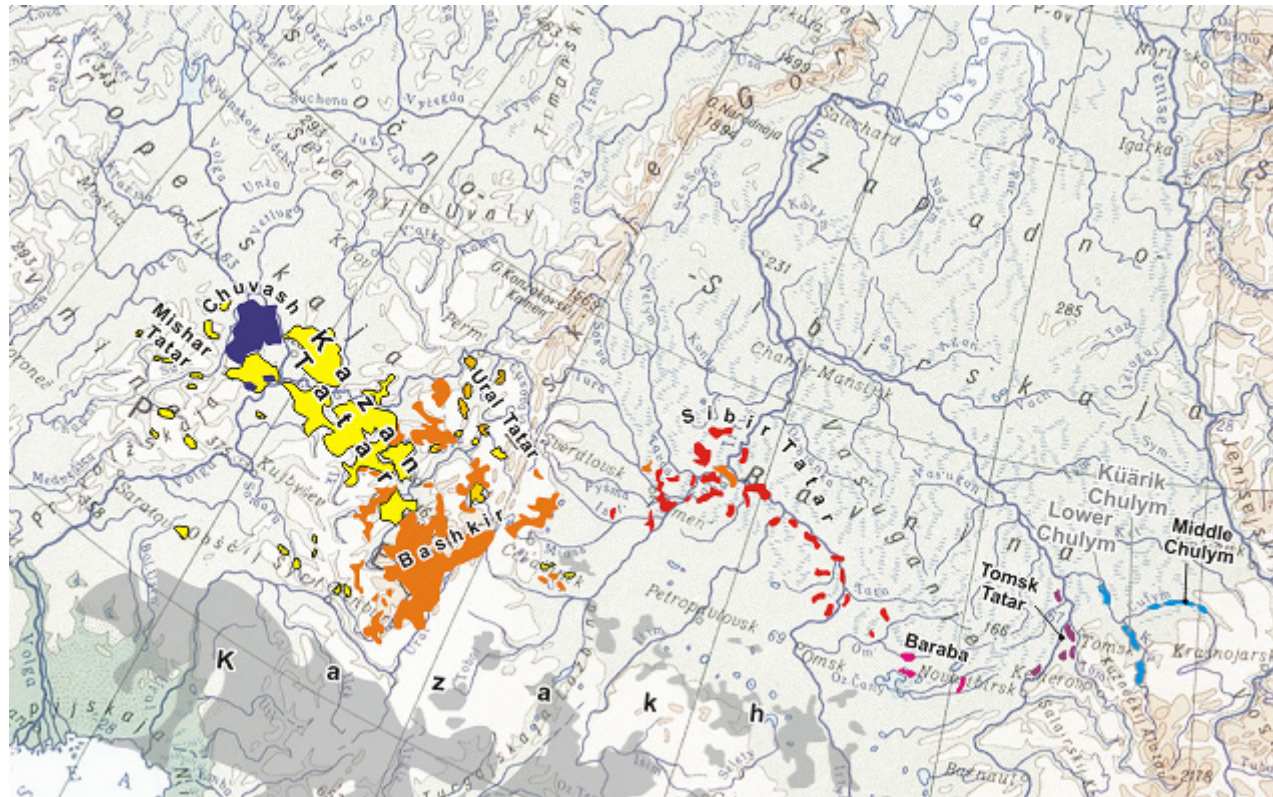
*He found them near the large river named Etil [= supposedly, Ak-Etil or Belaya, the main river of Bashkortostan]... And to everything he wanted to tell them, they listened carefully, for their language was entirely Hungarian, and they understood each other... The Tatar people live near them. But the Tatars, when waging a war on them, could not overcome them, on the contrary, they were defeated in the first battle... In that country, the aforementioned friar found the Tatars and the messenger of their lord, who spoke Hungarian, Russian, Cuman, Teutonian, Saracyn [=Arabic], and Tatar [and who said that behind the country of Tatars there were the "big-headed" people who wanted to start a war, perhaps the oncoming Mongols who must have reached West Siberia after 1207].*

*[Relatio fratris Ricardi, De facto Ungarie Magne a fratre Ricardo invento tempore domini Gregorii pape noni (On the existence of Magna Hungaria as related by Friar Ricardus), quoted from [a translation by S.A. Anninskiy \(1940\)](#)]*

This implies that the unusual phonological features in Bashkir could in fact have been the result of Tatar-Hungarian intermingling, when the local South Mansi and Majar tribes (=usually *Magyar* in Hungarian spelling) switched to Kimak-Kypchak-Tatar languages.

The interaction between Proto-Kazan-Bashkir and Proto-Hungarian had probably begun very early on, as implied by the very fact that the Hungarian expulsion from their homeland occurred as early as c. 830 AD, supposedly being caused by the warfare with the arriving Kimak tribes.

The interaction between the remaining Proto-Hungarians and the Bashkirs must have continued during the rise of the Golden Horde in the 14th century, when Turkic and Mongolian languages acquired significant importance in the region.





The distribution of the Kazan, Mishar, Ural, Bashkir, Sibir, Baraba, Tomsk Tatars and the nearby located Bulgaro-Turkic ethnicities  
[based on the *Atlas narodov mira (The Atlas of the Peoples of the World)*, Moscow (1964)]

### **On the Kazan-Bashkir interaction**

The glottochronological dates for Bashkir and Kazan Tatar predict a very recent physical separation — actually, only as late as the 18th century. Before that period, Bashkir and Kazan Tatar must have supposedly formed one single language.

Even if that date is exaggerated or results from a glottochronological error, Ibn-Fadlan's *al-Bashkird* people can hardly be directly equated with the speakers of the ancestors of modern Bashkirs of Bashkortostan. Linguistically, the *al-Bashkird* language must rather have been an early predecessor of Kazan Tatar, Bashkir and other local Kimak-Kypchak-Tatar languages.

But why do the languages known to exist rather separately for 1000 years, presently turn out to be so close to each other?

The reason is the lack of any natural geographic border between Bashkirs and Kazan Tatars, so as the map above shows, the mutual contacts never ceased and the two ethnicities must form a dialectal continuum.

Additionally, there was a long history of *Kazan Tatar, Mishar, Russian, Mari, etc. immigration to the Urals* and Bashkortostan that must have led to secondary linguistic exchange. There were various reasons for this movement, however one of the most significant was the strictness of feudal laws in Tsarist Russia and certain freedoms that Bashkirs were granted ever since their voluntary joining of the Moscovy in 1557. Consequently, Bashkir was probably continuously contaminated by Kazan Tatar, Russian and probably, to a much lesser extent, by Kazakh. The western, southern and standard (literary) dialect of Bashkir were particularly affected, with the eastern dialect being further located and less transformed by any external influence.

The immigration of Kazan Tatars into the Urals is also supported by the existence of a Ural dialect of (Kazan) Tatar or simply **Ural Tatar**. A researcher who studied these Ural Tatars (presumably before WWII) said that they claimed to have arrived in the Urals 500-600 years ago from the Volga, and seemed to be almost ethnographically indistinguishable from the Kazan Tatars. The Tatar immigration could have continued throughout the 18th-19th century because of the formation of metalworking industry attracting new workers to the Urals. [Sarmanajeva D.M., *Dialektnyje osobennosti yazyka sredneural'skikh tatar (The dialectal characteristics of the Middle Ural Tatars)*, dissertation, Kazan (1950)]

## **Conclusion:**

Accordingly, the present-day (Standard Literary) Bashkir and Kazan Tatar can be viewed almost as two varieties of the same language with a high level of mutual intelligibility. Naturally, when two languages are that close, the glottochronological principles imply that their separation should be very recent, obviously occurring already after the Mongol invasion of the 13th century.

The mutual proximity was even further strengthened by the Kazan Tatar immigration to the Southern Ural area resulting in secondary language contacts, which makes Kazan Tatar, Ural Tatar and Bashkir look and sound closer to each other than they are actually supposed to be historically judging by more than the 1000-year-long presence of the Bashkirs near the Southern Urals and the Kazan Tatars near the Volga River.

The odd Bashkir phonology can most likely be explained by the presence of several unknown substrata in the Southern Urals, such as South Mansi, or Proto-Hungarian, or western Samoyedic, or Bulgaric.

## **On the origins of Nogai**

### **Contesting Kazakh-Nogai direct genetic unity**

Much discussion has gone into contesting the direct Kazakh-Nogai genetic unity, which people of Kazakh and Nogai descent sometimes take for granted.

The theory was advanced by Baskakov in the 1950's through the 1980's, who was actually an expert in Nogai and published a Nogai dictionary in the 1940's. Indeed where there is the smoke, there is usually fire: as a matter of fact, there *are* certain features that indicate particular proximity of Nogai to Kazakh, whereas both languages share good mutually intelligibility.

However, the problem is not as simple as it seems. Most of the arguments against this hypothesis have already been expounded in the table for the Kimak languages, but we can add some more. The main criticism of all the Baskakov's hypotheses is that he was unable to differentiate between shared retentions and innovations, so most of his taxonomic suggestions were based merely on a few superficial phonetic and morphological shared features, not necessarily innovative ones.

In most of his works, namely [Baskakov, N.A., *Sovremennyye kypchakskije yazyki (The modern Kypchak languages)*, Nukus (1987)], [N. A. Baskakov, *Vvedeniye v izuchenije tyurkskikh jazykov (An introduction into the study of Turkic languages)*, Moscow (1969)], [*Ocherki istorii funktsionalnogo razvitiya tyurkskikh jazykov (The historical essays of Turkic languages functional development)*, Ashgabad, (1988)], which tend to repeat the same early content, Baskakov rather explicitly cites the following features for the Nogai-Kazakh subgrouping:

(1) the *ch* > *sh* mutation, as in Turkic \**kach-* > Nogai, Kazakh *kash-* "run away", Great-Steppe, Altay \**chach* > Nogai, Kazakh *sach* "hair";

(2) the *sh* > *s* mutation, as in Turkic \**qışh* > Nogai, Kazakh *qis* "winter", \**tash* > Nogai, Kazakh *tas* "stone";

However, similar changes are also present in Sibir Tatar, cf. Sibir Tatar *tas* "stone", *tsats* "hair", and Bashkir *säs* "hair";

**Note:** By *Sibir (Siberian) Tatar* we always understand "Tobol-Irtysh Tatar", whereas Baraba and Tomsk are seen as separate entities.

(3) The occasional retention of the "heavy" (fortified) consonant harmony, cf.

Nogai *qördiN be?* "did you see?" and Kazakh *Sen kinoga barasın ba?* "Are you going to the movies?"

However, this feature is also found in the 19th century's Baraba Tatar recorded by Radlov, cf. *Kildi ba?* "Did he come?" and, of course, in Kyrgyz *Keldi bi?* "Did he come?";

By the same token, we have Nogai accusative *-ni, -di, -ti, -ni, -di, -ti*, Kazakh *-ni, -di, -ti, -ni, -di, -ti*, however similarly, Baraba *-ni, -di, -ti, -ni, -di, -ti*, Bashkir, Kyrgyz *-ni, -ni, -nu, -nü, -ti, -ti, -tu, -tü*.

It should also be explained that, in any case, Kazakh is "heavier" than Nogai, which in other cases prefers the light western consonantism with lenition, e.g. Nogai *tas-lar*, as opposed to Kazakh *tas-tar* "stones".

(4) The usage of *-et-a-Gan* participle. Cf. Nogai *kel-et-a-Gan* "the coming one" and Kazakh *-atın / -etin*, etc. Not only these suffixes have different phonological shape in Nogai and Kazakh, they are also widely distributed among the Kimak languages as well, cf. Baraba *yör-ätiGän* "the usually walking one", Sibir Tatar *par-atıGan keshe* "a walking man";

And that is about all Baskakov mentions concerning the relationship of Nogai and Kazakh. So at this point, it seems that the *sh* > *s* and *ch* > *sh* mutation is the only typical common Nogai-Kazakh feature that is difficult to deal with.

We can also add a few of our own possible shared features and explain why they fail to correspond to the notion of a common proto-state:

(5) Nogai *-men* for instrumental case, as in Kazakh *at-pen* "with the horse", as opposed to Kimak \**belen*. However, this feature is not exclusive, and it is also present in Sibir Tatar, cf. Sibir Tatar *at-man* "with the horse". The usage of \**menen* or harmonically similar words can also be found in the southern dialect of Kazakh and Kyrgyz, e.g. *siz menen* "with you", Bashkir *menän*, Baraba Tatar *mīnan, mīna, ma:n*. As a result, this feature is hardly unique and is probably part of the local Sprachbund, whereas the contraction of \**menen* to *men* is also present in Sibir and Baraba. Moreover, based on other evidence, it must even go back to Proto-Bulgaro-Turkic, so its taxonomic value is arguable.

(6) The use of the archaic question word *qalay* "how" instead of *\*nichek* as in Kazan Tatar, Kumyk, Sibir Tatar, Baraba Tatar. However, in the Kara Nogai dialect we in fact do have *neshik* "how?", therefore *qalay* may be an old retention in Ak Nogai.

(6) The usage of a very specific Perfect Tense, cf. Nogai *bar-ïp-pan* "I have gone there" and Kazakh *bar-ïp-pïn* "it turns out I went". However, a similar tense seems to exist in several Kimak languages, cf. Sibir Tatar *par-ïp-mïn* "I used to go", Baraba Tatar *al-ïp-mïn* "It turns out I took", therefore it may be a retention.

(7) The active usage of the *\*ROOT-ïp (-a) + yat-* construction expressing Present Continuous, as in Nogai *bar-a-yatïr-man* "I'm going" and Kazakh *bar-a-zhatïr-mïn* "I'm going", *kel-ïp-tur-at* "He's coming", *ok-up-zhat-at* "He's studying", etc. But this feature was also widely distributed in Baraba (*ROOT-ïp + yat-, tïr-, ôtir-, yör-, kal-, bil-, al-*) and, of course Kyrgyz, e.g. *bar-a-jata-bïz* "We're going" as well as many other eastern Turkic languages. Therefore, it may be an old retention that survived in Nogai in a single construction *-a + yat-*.

(8) The usage of a quite characteristic and typical *I-want-to* construction, cf. Nogai *Men onï kör-Gïm kel-edi* "I want to see him", Kazakh *bar-Gïm kel-edi* "I want to go", literally "*desire-my came*". However it also exists at least in Kyrgyz *ayt-kïm kel-et* "I want to say" and Sibir Tatar *par-Gï kel-eu* "to want to go", let alone the Kazan Tatar parallels, therefore it is hardly unique.

(9) The usage of the Nogai *yew* "to eat" along with *ashaw* of Kimak origin, whereas Kazakh has only *zhew*. However, this is an obvious archaism and it also seems to be used parallelly in Sibir Tatar *ashau, yeü* "to eat".

(10) The use of Nogai *yapïraq* "leaf" and *top(i)raq* "earth", as opposed to Kimak *\*yapraq, \*topraq*. Note that an older Baskakov's dictionary [*Nogayskij yazyk i yego dialekty (The Nogay language and its dialects)*, Baskakov. N.A., Moscow (1940)] in fact provides *topraq*, so we may assume that both variants, *topïraq* and *topraq*, could be used interchangeably in Nogai. Cf. Kazakh *zhalbïraq, topïraq*. However this is an evident retention as it is also preserved in Kyrgyz *zhalbïraq, topuraq*; Altay *d'albïraq*; Khakas *tobïrakh*, Kumyk *topuraq*.

On the other hand, the more or less unique and purely Kazakh grammatical features that *must* be there, if the two languages were directly related, are *not* shared with Nogai, cf. the following Kazakh features:

(1) Kazakh *maGan*, but Nogai *maGa* "to me" as in all the TL's;

(2) Kazakh *bar-mak-pïn*, but Nogai *bar-ayak-pan* "I have to go, I will go", as in other Kimak languages. The presence of this unique feature was noted by Baskakov.

(3) Kazakh *siz-der bar-dï-Nïz-dar*, Kyrgyz *siz-der bar-dï-Nïz-dar* "you (plural) came", but no such construction in Nogai.

By the same token, none of the typical shared Kyrgyz-Kazakh isolexemes and iso-collocations are present in Nogai, even though they should be there:

- (1) Kyrgyz *chöp*, Kazakh *shöp* "grass", but Nogai *ölên* (as in other Kimak languages);
- (2) Kyrgyz-Kazakh *öte* "very", but Nogai *bek* (as in other Kimak languages);
- (3) Kyrgyz-Kazakh *özen* "river", but Nogai *yilGa suw* (as in other Kimak languages);
- (4) Kyrgyz *birö*, Kazakh *bireu* "someone", but Nogai *kim de*;
- (5) Kyrgyz-Kazakh *bir närse* "something", but Nogai *bir zat, ne di*;

### **Nogai vocabulary**

The Swadesh-215 lexicostatistics of Nogai (added in 2013, unpublished) shows the following values:

81% for the Nogai / Sibir-Tatar relationship;

81% for Nogai /Bashkir;

82% for Nogai /Kazan Tatar;

81% for Nogai /Kyrgyz;

82 % for Nogai / Kazakh;

79% for Nogai / Karachay-Balkar

This evidently makes Nogai equidistant from any other Kimak or Kyrgyz-Kazakh languages, which is of little help in determining its taxonomic position. Nevertheless, this suggests that Nogai could have formed as a Kimak dialect that absorbed some Kazakh elements.

### **Conclusion:**

We have found no unique Nogai-Kazakh innovations, which demonstrates that, despite all the mutual intelligibility, *the Nogai and Kazakh languages are of slightly different historical descent*, and their apparent proximity is mostly based on shared archaisms and secondary contacts.

When Proto-Nogai advanced from the Southern Ural area and the Tobol-Ishim Steppe towards the Jaik River somewhere between the 9th and 15th centuries, it must have retained archaic features which are also present in the 19th century's Baraba, modern Sibir Tatar and Kazakh, even though a later secondary influence from the western Kazakh dialect cannot be completely excluded.

Cf. the following retentions: (1) the retention of *-b-* as in questions, e.g. Nogai *qördiN be?*, Kazakh *barasiN ba?*, Baraba *kildi ba?*; (2) the retention of *-ti*, *-di*, *-di*, *-ti* in the accusative case in Nogai, (Radlov's) Baraba, Kazakh; (3) the retention of the 1st person singular *-mîn* in Nogai *bar-a-man*, Baraba *al-a-mîn* (Radlov), Kazakh, Kyrgyzbar-*a-mîn*, Tyumen Sibir Tatar *pel-ê-men*, as opposed to Kazan Tatar *bar-a-m*.

The only really interesting one-of-a-kind feature is the shared phonetic mutation *ch > sh*, *sh > s* that is also partly present in the Sibir Tatar transitions (*ch > ts*, *sh > s*), and to some extent in the Bashkir *ch > s* transition and even in Turkmen *s > β* (interdental or alveolar).

**Note:** This feature might show that before the arrival of the Great-Steppe tribes, there existed *a common substrate in the Ishim-Tobol-Emba-Yaik area* that had a very specific way of lenitive sibilant pronunciation. Judging by the superficially similar transitions in Chuvash, cf. Turkic *\*chach* vs. Chuvash *s'üs'e* "hair" and Turkic *\*tash* vs. Chuvash *chol*, we may tentatively assume that this substrate might have possibly been of Bulgaric origin; or at least this possibility cannot be excluded.

In any case, a possible existence of this substrate has no direct bearing on the supposed Kazakh-Nogai unity, which was the point of the discussion above.

The cases of Uzbek-Uyghur, Bashkir, Nogai show that *in closely related languages, taxonomic conclusions cannot be based upon superficial phonetic similarity alone, since such features may result from a secondary mutual exchange with each other or a third language*. A presence of unique grammatical and lexical innovations is required instead.

Many doubts remain, however, and the exact prehistory of the Proto-Nogai dialect and its interaction with Kazakh remains unclear.

### **Karachay-Balkar, an atypical Kimak language**

Most features listed in the table above indicate that Karachay-Balkar (self-appellation: Qarachay-Malqar) also belongs to the Kimak languages. However, much evidence sets it apart as a distinctive and peculiar Kimak representative from the North Caucasus.

### **Karachay-Balkar phonology**

In most respects, Karachay-Balkar share the same typical innovations as other Kimak-Kypchak-Tatar languages, such as (see the table above):

- (1) a mixed *-Ga /-a* ending in the dative case;
- (2) the traces of an intervocalic sound in *baur* < *\*bawur* "liver", *süyek* "bone";
- (3) a typical Kimak suffix in *ur-luk* "seed";
- (4) the softened (lenitive) *-d-* > *-l-* transition as in *-juk-la-* "sleep", *-la* "the plural suffix".

However, certain other features set Karachay-Balkar apart from the typical representatives of the Kimak-Kypchak-Tatar subtaxon, such as:

- (1) the retention of */J-/*, */ch-/*; note that, as we have shown above, the initial *J-* / *ch-* is supposed to be present in Proto-Turkic;
- (2) the retention of */t-/* in *tört*;
- (3) the retention of the *-Gaq* suffix, as well as a few phonological innovations probably from the Circassian-Kabardian substratum;
- (4) the loss of *-r* in *-lar* / *-ler* ;

### **Karachay-Balkar grammar**

Among the most typical Kimak-Kypchak-Tatar grammatical features, one could name the following:

- (1) the use of the future tense with the *-rik*, *-nik*, *-lik* suffix, apparently akin to the Oghuz and Tatar *-aJak*, *-eJek*;
- (2) the use of *tüyül* instead of *emes*;

Among peculiar features, there is the formation of the Present Tense in Karachay-Balkar using the *-dir-*suffix, which is also found in Altay-Sayan and Sakha:

*ROOT + -a/-e + tur + personal ending = Present Continuous*

### **Karachay-Balkar vocabulary**

Lexically, Karachay-Balkar is almost equidistant from other languages of the Great Steppe: **78%** from Tatar-Bashkir and about **78%** from Kyrgyz-Kazakh (most likely due to the high retention of archaisms in Kazakh-Kyrgyz); 75-76% from Uzbek-Uyghur, 69% from Turkmen, 65% from Standard Altay and Khakas (Swadesh-215).

The lexicostatistical research suggest the early separation of Karachay-Balkar from the Kimak stem, basically occurring at the same period as the Kyrgyz-Kazakh, which is approximately consistent with the existence of the Kimak Kaganate unity near the Irtysh. The glottochronological separation date is about 730 AD, but this figure may be set too low, considering that the Circassian-Kabardian influence was not taken into consideration.

Circassian and Kabardian are the two neighboring languages of Northwest Caucasian stock, which are distantly related to each other. Their presence seems to have resulted in certain Caucasian borrowings into the basic Karachay-Balkar vocabulary. At least the following Circassian words were found in Swadesh-200 (1%):

Karachay-Balkar *gakkī*, Circassian *qanqa* "egg";

Karachay-Balkar *gokka*, Circassian *qeGeG*, Kabardian *GaGe* "flower";

### **Karachay-Balkar history**

The early history of Karachay-Balkar is poorly understood. A likely date for the Proto-Karachay-Balkar arrival in the Northern Caucasus is circa 1000-1050 AD, when the Kypchak-Cuman-Polovtsian tribes began to infiltrate into the Pontic steppes and finally appeared near the Kievan Rus. However, historically, the Karachay-Balkar people are only attested since the Mongol invasion or even centuries later.



### **Conclusions:**

The lexical differences set Karachay-Balkar aside from other representatives of the Kimak-Kypchak-Tatar subtaxon, however the presence of certain grammatical and some of the phonological innovations is quite in accordance with the Kimak origins of Karachay-Balkar. Generally, we should assume an early separation of Karachay-Balkar from the Kimak stem, that occurred somewhere circa 800-900 AD. This separation was probably unconnected with the Mongol invasion and the later expansion of dialects of the Golden Horde, but occurred a few centuries earlier when Proto-Karachay-Balkar tribes moved towards the North Caucasus.

After settling in the Caucasus, Proto-Karachay-Balkar was to some extent affected by its North Caucasian neighbors, whose influence is now evident at least in the basic vocabulary.

### **The Oghuz-Seljuk subtaxon**

#### **Oghuz is still a valid subtaxon**

The Oghuz-Seljuk subtaxon (traditionally named just *Oghuz*) includes at least the following western Turkic languages:

- (1) the Turkmen dialects, namely Teke, Yomud, Ersarin, Saryn, Saryq, Chovdur, Trukhmen;
- (2) Azeri, Qashqai, Turkish and Gagauz.

The taxon is characterized by a number of distinctive features described below.

#### **Oghuz-Seljuk phonology**

In phonology, the Oghuz-Seljuk subtaxon is marked by the famous *Oghuz voicing of initial consonants* ( $t- > d-$ ,  $k- > g$ ).

Note, however, that the Oghuz voicing has never been conclusive or comprehensive — as it has been shown (at least) by A. Scherbak (1970) [(cited in detail by Starostin in *The Altaic Problem and the Origins of the Japanese Language* (1991))], many words in Turkmen, Turkish, and Azeri preserve the word-initial  $k-$  or  $t-$ , a trait that may go back to the Oghuz proto-state or that may have developed because of the Karakahnid and Great-Steppe influence, e.g. Turkmen *towuq*, Azeri *toyug*, Turkish *tavuk* "hen"; Turkmen *kim*, Azeri *kim*, Turkish *kim* "who", etc.

Moreover, note that many other Turkic languages exhibit temporary intervocalic voicing, e.g. Kyrgyz /*maGa gelseN*/ "if you come to me" (written as *maga kelseN*).

Also see the phonological comparison with Orkhon-Karakhanid below.

### **Oghuz-Seljuk grammar**

Several shared Oghuz-Seljuk innovations can be found in grammar, such as:

(1) The *full transition of -ga/-ge, -ka/-ke into -a/-e in the dative case*;

(2) The *loss of m-/b- in the 1st person plural -bis / -mis* verbal ending marker, hence Turkmen *-is*, Turkish *-iz* and Azeri *-ik* (where the original *-iz* has been further replaced by the past tense suffix);

(3) The frequent use of *the synthetic Present Continuous Tense with -yor-*, cf. Turkmen *-yar-*, Azeri *-yur-*, *-ir-*, Turkish *-yor-*, apparently originating from the verb *yürü-* "to walk, go" and the syntactic constructions similar to those used in the Great Steppe languages, e.g. originally Proto-Oghuz *\*bar-ıp jürü-r* or *\*bar-ıp jörü-r* "he is leaving", but presently Turkish *var-ı-yor* "he is arriving". Cf. Turkmen *okap yör* "he is still learning", *gezip yör* "he is walking around". The verb *jürü-/yürü-* "to go" is used here essentially in the same way as in Kyrgyz and Kazakh, which implies that the construction may be a Great Steppe borrowing created circa 600-700 AD during the contacts with the Great Steppe languages near the Zaisan Passage.

### **Oghuz-Seljuk vocabulary**

A few examples of the Oghuz-Seljuk innovative isolexemes are listed below. These have mostly been found in Swadesh-215 and they all belong to the basic vocabulary.

**Note:** Please note again the difference between a *cognate* and *isolexeme*. Even though some of the cognates may also be known in some other languages or some borderline dialects having a different meaning, an isolexeme in this particular phonological shape and this particular meaning exist only in this specific language branch and territory.

Some of the words below may also be occasionally found in languages that were in contact with the Oghuz (Crimean Tatar, Crimean Karaim, Kumyk, northern Uzbek dialects, Karakalpak etc.) where they may constitute Oghuz borrowings;

(1) Turkish *bura-(da)*, Azeri *bura-(da)*, Turkmen *bäri* presumably from *bu yer* (or less likely *bu ara*) "this place (span)", also cf. Kazan Tatar *bire-dê* "here", which shows that this word seems to have been borrowed into Kimak-Kypchak-Tatar, as opposed to *\*munda* and *\*bu yerde* in other western Turkic languages;

(2) Turkish *nere-(de)*, Turkmen *nire-(de)* "where" from *ne yer-de* (or less likely *ne ara-da*) "which place (span)";

(3) Turkish *çok*, Azeri *chox* "many, very", Turkmen *choq* "a crowd", as opposed to *toköp* in other western Turkic language;

(4) Azeri *chaga*, Turkmen *chaga* "child", Turkish *chaga* "baby", as well as Turkish *choJuk* ("child" < "piglet"), as opposed to *bala* in most other Turkic languages;

(5) Turkish *kök*, Azeri *kök*, Turkmen *kök* "root"; not found in other Turkic (?); apparently a curious retention from the Bulgaro-Turkic level, cf. Chuvash *kâk kâkla* "to uproot the tree stumps". It is also found in Kazakh in the meaning "roots, pedigree" (apparently from Oghuz), and in Karakhanid.

(6) Turkish *ada*, Azeri *ada*, Turkmen *ada* "island"; acc. to Severtyan's Dictionary may also be found in some languages in contact with Oghuz (Crimean Tatar, Crimean Karaim, Uzbek dialects, etc)

(7) Turkish *çek-mek*, Azeri *chäk-mäk*, Turkmen *çek-mek* "to pull", as opposed to the variants of the *tart-* root in most other Bulgaro-Turkic languages.

(8) Turkmen *kütek*, Azeri *küt*, Turkish *küt* "dull (as of a knife)", as opposed to *\*otmes*, *\*maka*, etc. in other TL's.

(9) Turkish *köpek*, Azeri *köpäk*, Turkmen *köpek* "dog", as opposed to a more archaic *it* in other TL's, which is also used in Turkish and Azeri but less frequently. Essentially, *\*köpek* seems to be an Oghuz word, though it can also be found in other borderline TL's where it is much less common;

(10) Turkish *geniş*, Azeri *genish*, Turkmen *ginish* "wide" with the *-sh* suffix.

[Besides these languages, the Severtyan's dictionary apparently incorrectly cites Kyrgyz, where *keNish* means "widening" [see Yudakhin's dictionary of Kyrgyz], and Karakalpak, where "wide" is naturally *keN* as in most other TL's, such as Tatar, Bashkir, Karachay, Kazakh, Kyrgyz, Karakalpak, Uzbek, Uyghur];

(11) Turkish *üfle-mek*, Azeri *üflä-mäk*, Turkmen *üfle-mek* "to blow (at something, e.g. a candle)";

(12) Turkish *dön-mek*, Azeri *dön-mäk* "turn (right, left, back)", Turkmen *dön-mek* "return, turn back". Cf. also Tatar *tün-* "to turn over (upside down)" and probably other similar words in Kimak-Kypchak-Tatar languages but with semantical differences. In any case, the word seems to be originally Oghuz;

(13) Turkish *saG*, Azeri *saG*, Turkmen *saG* "right (side)". Acc. to Clauson, from the original meaning "healthy" connected with the purity of right-handedness in Islam, which seems a reasonable etymology;

(14) Turkish *günes*, Azeri *günäsh*, Turkmen *günes* "sunny (side), sun", as opposed to just *gün* in most other Turkic languages, though the latter is used in Oghuz-Seljuk just as well;

(15) Turkish *düz*, Azeri *düz*, Turkmen *düz* "smooth", as opposed to *\*tegiz* in most languages of the Great Steppe. The lexeme is also found in Altay-Sayan languages in the same meaning, albeit this is perhaps coincidental;

(16) Turkish *kurt*, Azeri *kurd*, Turkmen *gurt*, *möjek* "wolf", apparently, originally pejorative from "a bug, parasite", that is "a parasite that kills the sheep"; the lexeme may also be a folksy Turkic elaboration of the Persian *gurg* "wolf"; it was mentioned by Mahmud al-Kashgari c. 1073 as an Oghuz word; whereas most other Turkic languages use a more archaic lexeme *\*böre*;

(17) Turkish *geche*, Azeri *gechä*, Turkmen *giye* "night". An archaism, judging by the fact that it exists in Chuvash as *kas'*, which shows that this might have been the original way to say "night", probably subsequently displaced by *tün* in most Turkic languages after their separation from Bulgaric. It is also inconsistently found in Karachay, Crimean Tatar (most likely from Ottoman Turkish), Uzbek and Salar, which seems to confirm that this word is an archaic retention;

(18) Turkish *dösh* (*colloq.*), Azeri *dösh*, Turkmen *dösh* "breast", as opposed to *\*emchek* in most other Turkic languages; on the other hand, also cf. Kyrgyz *tösh* "breastbone, sternum", Kazakh *tös* "breast" etc., therefore probably an archaism;

As you can see, there exist multiple Oghuz-Seljuk isolexemes.

The average lexical distance in Swadesh-215 from Oghuz-Seljuk to Great-Steppe is only about 69%, making them rather mutually unintelligible in real speech, whereas the distance to any other major branches is even greater, clearly setting Oghuz-Seljuk aside from other Turkic languages.

## Oghuz history and geography

The Oghuz people first appear in history after 605 or 630 AD [see S.G. Klyashtornyi, *Stepnyye imperii: rozhdeniye, triumfi i gibel (The Steppe Empires: birth, triumph and disintegration)*, Saint Petersburg (2005) ]. They are clearly mentioned in the Orkhon inscriptions circa 720, which makes them, along with *Qirgiz* and *Türük*, one of the oldest historically attested Turkic clan confederacies.

In the Orkhon inscriptions, they are described as the *Toquz Oghuz* tribal union that waged war with the *Tür(ü)ks*, but was finally conquered and subjugated by them. Therefore, a clear ethnological difference between the *Türük* and the *Oghuz* tribes has been evident starting from the earliest historical records, which implies that the *Oghuz* tribal confederacy must have formed as a distinct linguistic and ethnographic entity at least a few centuries before their first attestation, that is *before* 600 AD.

Outside the famous *Toquz Oghuz* "The Nine Oghuzes", there existed other ethnonyms of the same structure, such as *Seqiz Oghuz* "The Eight Oghuz" [mentioned in the *El Etmish Bilge Kagan* inscription (759)], *Otuz Tatar* "The Thirty Tatars" [idem], *Üch Qarluq* "The Three Karlucs" [idem], etc. Therefore, the number before the ethnonym could easily change depending on political circumstances, and apparently just *denoted the number of clan units forming a tribal confederacy*. Continually mentioning this number before the clan name must have been important from the military and diplomatic point of view, because it showed how many tribal units participated in a given conflict or how potent and influential they could be.

Being first historically attested in the 720's in Mongolia as enemies of the *Tür(ü)ks* and vaguely located somewhere in the vicinity of Kyrgyz and Tatar tribes, the Oghuz tribes have rather uncertain geographic origins. However, the most plausible interpretation of their earliest position points *somewhere near Dzungaria*.

The typical southern location of the Oghuz beyond the Tian Shan may immediately imply their geographic connection to the southern Orkhon-Karakhanid branch of the Turkic languages. This is indeed easily corroborated by such isolexemes as *deGilfound* (outside the Kimak subgroup) only in Old Uyghur, and occasionally in Chagatai, as well as by other features described below.

However, a definite geographical attestation, such as provided by Ibn-Fadlan in 921, pinpoints the Oghuz tribes already in the Aral-Caspian region, when they were partly Islamized.

*Ibn al-Athir, an Arab historian, explained that the Oghuz tribes had moved to the Syr-Darya River in Transoxania during the period of the caliph Al-Mahdi (775-785)*, therefore their westward movement, probably along the northern Tian Shan, must have been rather quick on historical scale.

It is plausible to assume that the Transoxanian Oghuz tribes were concentrated along the shores of the Aral Sea, where they naturally must have engaged in fishing to supplement their diet with maritime products. As a matter of fact, there are certain (albeit rather inconclusive) historical,

archaeological and journalist reports of tomb ruins and towns, such as *Yangi Kent*, found in the vicinity of the Aral Sea. Being dated to the first millennium, they probably belonged to the Oghuz, but any archaeological issues are outside the main scope of this article and cannot be discussed here at much length.

### **Conclusion:**

The Oghuz-Seljuk subtaxon constitutes a linguistically valid unity that must have originally inhabited an area near Dzungaria and Lake Zaysan from 600 (or earlier) at least until the 9th century. Circa 780 AD, they (partly?) migrated towards the Syr-Darya River and then to the Aral Sea, apparently moving along the northern track of the Silk Road near the foothills of the Tian Shan, which is the shortest and most suitable route that avoids arid areas of central Kazakhstan. By the 920's, the Oghuz people were clearly described in the region located between the Aral and Caspian Sea by Arab traveler Ibn Fadlan.

### **Seljuk as a subtaxon of Oghuz**

Secondly, there are certain innovative features that separate the Seljuk languages, such as Turkish, Gagauz and Azeri, from the Turkmen dialects, which makes it necessary to differentiate the Seljuk subtaxon from the rest of the Oghuz languages.

As a result, we will normally use the term *Oghuz-Seljuk* instead of just Oghuz to stress the composite nature of this subtaxon.

### **Seljuk vocabulary**

The following isolexemes in Swadesh-215 are absent from Turkmen, making Turkish and Azeri particularly close to each other. The comparison with Turkmen was made using a dictionary of the Standard (Literary) Turkmen [*Kratkij russko-turkmenskij slovar*, Editors-in-Chief: M. Khazmayev, S. Altayev; Ashgabad (1968)], so any particularities of other Turkmen dialects were not taken into consideration.

- (1) *say-mak* "to count (numbers)", cf. Turkmen *sana-mak* "to count" and *say-mak* "to believe, think";
- (2) *sil-mek* "to wipe (dust)", cf. Turkmen *süpür-mek* of the same meaning;
- (3) *bura-da* "here (locative)", a phonological innovation, as opposed to Standard Turkmen *bu yerde, mında, shu tayda*, etc;
- (4) *ora-da* "there (locative)", as opposed to Standard Turkmen *ol yerde, ol tayda*;

(5) Turkish *çok*, Azeri *çok* "much, many; very", an innovation, as opposed to *köp* in Turkmen and most languages of the Great Steppe Sprachbund;

(6) *düş-ün-mek* "think", a semantic innovation, as opposed to "understand, know" in Turkmen and other languages of the Great Steppe Sprachbund;

(7) *vur-mak* "hit", with the innovative /v-/, as opposed to \**ur-* in Turkmen and most Turkic languages;

(8) Turkish *ol-mak*, Azeri *ol-mâq* "to be", as opposed to *bol-* in Turkmen and most languages of the Great Steppe; a rarely occurring and rather irregular phonological innovation also present in Turkish *ile*, Azeri *ilä* versus Turkmen *bilen* "with (someone)"

(9) Turkish *var-mak*, Azeri *var-mak* "to arrive", a semantic innovation, as opposed to the Turkmen *bar-mak* "to go, walk, visit" as in other Turkic languages; actually, *bar-* is a very typical Turkic verb with the meaning "to go (somewhere)"; the original meaning of the Seljuk verb *var-* is retained in Turkish in the imperative *Var!* "Go; do as you wish!"; it was for instance frequently attested in this way in an 18th century's Turkish-English phrasebook when giving directions to a boy, a salesman at an Ottoman market, etc.;

(10) Turkish *ait*, Azeri *aid* "belonging to", a semantic innovation; the verb *ayt-mak* "to speak, talk" is very common in most languages of the Great Steppe Sprachbund, including Turkmen, but acquired a different unrelated meaning in Proto-Seljuk;

(11) Turkish *on-lar*, Azeri *on-lar* "they", but simply *o-lar* in most other languages from Turkmen to Tuvan;

(12) Turkish *kısa*, Azeri *kısa* "short", but *qısqa* in most other languages from Turkmen to Tuvan;

(13) Turkish *kadın*, Azeri *qadın* "woman", probably an old retention, instead of *heley, ayal* (from Arabic) in Turkmen and many languages of the Great Steppe;

(14) *bağirsak* "intestine (gut)", evidently formed from *bağır* "liver", cf. *ichege* in most Turkic languages including Turkmen; this word is unlikely to be a Seljuk innovation taken that it can also be found in Bashkir and some other Kimak languages with slightly different meanings, acc. to Sevortyan's Dictionary, even though there is hardly any direct confirmation from modern dictionaries of these languages; also cf. Chuvash *pirshâ-lâx* "intestines, guts"; probably an Oghuz partial innovation subsequently lost in Turkmen;

(16) Turkish *orman*, Azeri *orman* (poetic), usually *meshä* "forest" versus Turkmentokay, *zheNNeI*; The word is actually found in many Turkic languages of the Great-Steppe (Kazan Tatar, Bashkir, Nogai, Kazakh, Uzbek, Uyghur, moreover cf. Chuvash *vârman* "forest" where it seems to be borrowed from Kazan Tatar); judging from the relative scarcity of forests near the Dzungaria Desert, the word *orman* might have been a

borrowing from Proto-Great-Steppe into Proto-Oghuz with a subsequent loss in Standard Turkmen; alternatively, it could be a Turkic or even Bulgaro-Turkic retention;

(17) Turkish *uyu-mak* "to sleep", Azeri *uyu-mäk* "to fall asleep", cf. Turkmen *ukla-mak*, Uzbek *uxla-moq*, Uyghur *uxli-maq* "to sleep"; an Oghuz retention subsequently lost in Turkmen;

(18) Gagauz *ev*, Turkish *ev*, Azeri *ev* "home", as opposed to *öy* in most languages of the Great Steppe Sprachbund; probably an Oghuz retention subsequently lost in Turkmen;

(19) Turkish, Azeri *her shey* "everything" from Persian, cf. Turkmen *herşi* "every", but *hemme, bari* "everything"; a borrowing into Proto-Seljuk;

Lexicostatistically, there is merely a poor relatedness of 74% between Turkish and Standard Turkmen and 78% between Azeri and Standard Turkmen. By contrast, there is a much better Turkish-to-Azeri lexical overlapping of 86% (Swadesh-215, Persian and Arabic borrowings excluded).

The Turkmen subtaxon is about 5% closer to the languages of the Great Steppe than to Turkish-Azeri, cf. 73% Turkmen / Chagatai-Kyrgyz-Kazakh; 70% Turkmen / Kimak; 67% Seljuk / Chagatai-Kyrgyz-Kazakh; 66% Seljuk / Kimak. Therefore the Turkmen subtaxon seems to be more affected by Persian and the languages of the Great Steppe because of the interaction with Kazakh, Karakalpak, Nogai and Uzbek, whereas the Seljuk subtaxon seems to retain more archaic features because of its early separation.

### **Seljuk history and geography**

The split of the Seljuk clan from the Oghuz tribal confederacy in 985 resulted in an early diversification of the Aral Oghuz tribes into the Turkmen and Seljuk subbranch. The subsequent formation of the Great Seljuk Empire by Tughril Bek in 1037 is well-known from historical records.

### **Conclusion:**

The Aral-Caspian position of the Turkmen Oghuz suggests that *the Transoxanian Oghuz language must have existed in a close contact with the languages of the Great Steppe* from the 8th century onward, and was therefore affected by Nogai, Kazakh, Karakalpak and Uzbek, thus acquiring



certain features typical of the Great Steppe Sprachbund. As a result, Turkmen presently forms a separate subtaxon within the Oghuz-Seljuk branch and includes a variety of Turkmen language-dialects, which are rather superficially described in the Turkological literature.

On the other hand, the *Proto-Seljuk* language spoken in the Great Seljuk Empire led to the rise of Ottoman Turkish, Gagauz, Azeri, Qashqai and presumably other distinct Seljuk dialects in Persia and Anatolia.

### **Oghuz-Seljuk is indirectly related to Orkhon-Karakhanid**

At first glance, the Oghuz-Seljuk languages seem to share a number of linguistic features with Orkhon and Karakhanid languages. However we need to find specific evidence clearly substantiating the direct descent of Oghuz-Seljuk from Orkhon-Karakhanid, so we have to study the Oghuz-Karakhanid relation in more detail.

Naturally, some of the Orkhon-Karakhanid features are also found in modern Uyghur and Uzbek, which inherited certain traits from Karakhanid, so instances from these languages may also be listed below, even though they presently belong to the Great-Steppe subtaxon.

### **Oghuz and Karakhanid phonology**

In phonology, Oghuz and Karakhanid share the following features:

(1) the presence of *the intervocalic -G- and the word-final -G*, as in Turkmen *baGir* "liver", *aGir* "heavy"; Uyghur *beGir*, *eGir*; Uzbek —, *oGir*; Karakhanid *baGir*, *aGir*; Turkish, Azeri, Turkmen *daG* "mountain", Uzbek, Uyghur, Karakhanid *taG*; this may be either an archaism or innovation;

(2) a typical sonorization pattern as in *\*sekkiz*, *\*doquz*, as opposed to the Kimak-Kypchak-Tatar *\*segiz*, *toGuz*; rather an archaism

(3) the retention of *the nasal -N-* or its modification as in Azeri *sümük*, Turkmen *süNk*, Uyghur *söNek*, Orkhon Old Turkic, Karakhanid *söNük* "bone"; probably an archaism;

(4) the lenition of *-d-, -t-, -l-* > *-l-* as in *-lar*, *-ler*; this feature could rather be called *the light Turkic consonantism*. It is also shared by Kimak languages, especially west of East Bashkir, Baraba, etc. and other areas outside of West Siberia. This feature is most likely an old Orkhon-Oghuz-Karakhanid innovation that spread to Kimak from Oghuz when they must have been in contact near Lake Zaisan (see above);

On the other hand, the Oghuz-Seljuk languages exhibit certain phonological features which clearly differentiate them from Karakhanid and Old Turkic. Makhmud al-Kashgari's (1073), for instance, cited over 200 Oghuz-specific words and a number of classical phonological Oghuz

mutations. These *classical Oghuz phonological mutations*, present as early as the 11th century, allowed him to distinguish the medieval Oghuz language-dialect from Karakhanid:

- (1) **m- > b-** as in Oghuz "I" (the *ben* pronoun is presently found mostly just in Turkish);
- (2) **t- > d-** as in Oghuz "camel";
- (3) **w- > v-** as in Oghuz "hunt";
- (4) **-G- > -0-** as in Karakhanid vs. Oghuz "throat", Karakhanid vs. Oghuz "going, gone";
- (5) **-D- > -y-** as in Oghuz <äyig> "bear", "birch" with the loss of -ð- as opposed to the Karakhanid, evidently because of the Great-Steppe influence where the same transition is inherited from an earlier Proto-Central level.

As a result, Al-Kashgari (1072) described Oghuz as a dialect *quite different* not only from Kypchak, but also from the "normal" and "pure" Turkic, which to him naturally was Karakhanid, implying there was a rather early differentiation between Oghuz and Karakhanid languages.

### **Oghuz, Karakhanid and Orkhon grammar**

Oghuz Seljuk, Old Uyghur, some Uzbek dialects, Karakhanid and Orkhon grammars are all characterized by the frequent use of *-miş-* in the audative mood. The *-miş-* suffix (1) can join nouns and adjectives, cf. the contracted form of *i-mish*; (2) it can be used as a perfect participle; (3) it can be used as a perfect tense suffix. The primary and the most usual function of *-miş-* in spoken Oghuz-Seljuk is to express astonishment and reported speech.

However, *-miş-* is not used in Standard Turkmen that uses *-a:n* in the perfect tense just as other languages of the Great Steppe.

The use of a *-miş-* cognate as the past tense suffix is also typical in Sakha where the suffix *-bit-, -bit-, -büt-, -but-, -püt-, -pit-, -püt-, -put-, -müt-, -mit-, -müt-, -mut-* is used to denote the perfect tense.

The usage of *-miş-* to express astonishment is also mentioned in Uzbek. Besides, even though *-miş-* is no longer used in modern Kimak-Kypchak-Tatar, it was used as past tense in Cuman-Polovtsian. It also seems to be sometimes found in Chagatai. But in any case, it must be an archaic morpheme surviving in Seljuk, Orkhon-Karakhanid and Yakutic.

The phonological and harmonical structure of *-miş-* suggests that its equivalent was Proto-Bulgarian *\*-bul-*, which implies that it might have originally formed from the verb *\*bol-* "to be" in the same way as composite tenses with the substantive, auxiliary verb *tend* "to be" are formed in many languages.

### **Oghuz, Karakhanid and Orkhon vocabulary**

Most of the Oghuz-Seljuk-specific words can in fact be explained from Karakhanid sources [see *Drevnetyurkskiy slovar (The Old Turkic dictionary)*, Editors: V.M Nadelyayev, D. M. Nasilov, et al., Leningrad (1969)]. Cf. the following examples:

(1) Oghuz *\*el* (hand), Karakhanid, Old Uyghur *eliG* (also found in Chuvash, Sakha, Yugur); this word is not shared by Uzbek, Uyghur, Kimak-Kypchak-Tatar;

(2) Oghuz-Seljuk *choq* "much, very", Karakhanid *choq* "much, very";

(3) Oghuz-Seljuk *kök* "root", Karakhanid *kök* "root";

(4) Oghuz-Seljuk *geche*, Karakhanid *keché*;

(5) Oghuz-Seljuk *dösh* "breast", Karakhanid *tösh*;

(6) Oghuz-Seljuk *chek-*, "to pull", Karakhanid *chek-* "to pull; tie";

(7) Oghuz-Seljuk *köpek*, Karakhanid *köpêk* "dog";

(8) Oghuz-Seljuk *günes* "sun", Old Uyghur (?) (attested in the *Irq Bitig*) *künêsh*;

(9) Oghuz-Seljuk *düz* "smooth", Orkhon Old Turkic, Karakhanid *tüz*;

(10) Seljuk *ev* "home", Karakhanid *ev*;

(11) Seljuk *uyu-* "to sleep", Karakhanid *uDî-*;

The retention of the many Orkhon-Karakhanid archaisms in Oghuz-Seljuk is evidently indicative of the Oghuz relatedness to the Orkhon-Karakhanid subtaxon at the lexical level.

### **Oghuz, Karakhanid and Orkhon history and geography**

Curiously, using certain historical records, S.G. Klyashtorniy describes the Toquz-Oghuz tribes as something that has naturally split off from the Uyghur tribal confederacy.

*In 605, [...] the Uyghur leader has taken his tribes to the Khangai Mountains [= in eastern Mongolia], where a separate group was created, known in Chinese historiographical sources as "the nine tribes". In the Orkhon inscriptions, this group was named Toquz-Oghuz.*

[*Stepnyye imperii: rozhdeniye, triumf i gibel (The Steppe Empires: birth, triumph and disintegration)*, Saint Petersburg (2005)].

Therefore, we may assume that *Oghuz* is nothing but a different pronunciation of *Uyghur*, which can easily be explained by the widespread usage of the liquid affricate in Mongolian (and most likely the nearby early Turkic languages and dialects), where /r/-/l/-/s/-/z/ are in some cases pronounced as mere allophones of the same phoneme. In other words, it is not even necessary to add any evidence from the Bulgaric languages, where the /z/ to /r/ mutation is compulsory, rather the local Khalkha Mongolian data provide enough substantiation, since the -z to -r mutation could have arisen either on the basis of incorrect Mongolic-based translations, transcriptions, reinterpretations, Sprachbund phonology, etc. In any case, the hypothesis that Oghuz and Uyghur may have originally been the same ethnonym seems quite plausible, albeit not clearly demonstrated.

In any case, the scanty historical records confirm that the earliest Oghuz tribes must have been located somewhere between the Tarim Basin, the Khangai Mountains and Dzungaria, probably near the Mongolian Altai and the Dzungarian Gobi.

Therefore, using this geographic perspective, we may conclude that Proto-Oghuz must have originally been a Dzungarian variety of Orkhon-Uyghur-Karakhanid, that had initially moved towards Mongolia but either stayed midway in Dzungaria or even *turned back again from Mongolia towards the Altai and / or Mongolian Altai Mountains*. This Proto-Oghuz backwave probably occurred by the 6th century AD during the initial rise of the Göktürk Kaganate. As a result, the Oghuz superstratum apparently traveled back through the Zaysan Passage towards the Irtysh river where it must have run into the Kyrgyz tribes, or the speakers of various Kyrgyz-Karluks dialects (see above [The relationship between Oghuz and Kimak](#)).

### **Conclusion:**

On one hand, the Orkhon-Karakhanid-Old-Uyghur features in Oghuz-Seljuk are remarkable and Oghuz seems to be rather clearly related to Karakhanid and Old Uyghur considering that it shares both archaic retentions and innovations, and even bears nearly the same name. Moreover, historical sources seem to vote for the split of Oghuz from Old Uyghur circa 605 AD.

On the other hand, the phonological changes in Oghuz, as compared to the Karakhanid of the 1070's, should have taken some glottochronological time to develop, and are probably consistent with about 500 years of separation, therefore we should conclude that Oghuz was *not a direct* offshoot of Karakhanid, but rather its sibling that had separated from the Old Uyghur stem circa 600 AD.

So we arrive at a conclusion that Oghuz was a different branch of Orkhon-Karakhanid dialects that *must have traveled a different geographic route from the Altai region without getting intermingled with the Kara-Khanid and Kara-Khoja dialects of the Tarim Basin*. As it has been described above, the only alternative route available was located *north of the Tian Shan Mountains*. And indeed, we do know from historical records that this route was explored by the Gökturks as early as 600-700s AD. We also know that the Oghuz tribes must have migrated from the Irtysh to the Syr-Darya River along this Silk Road somewhere circa 780 AD. Consequently, our linguistic analysis seems to confirm the historical evidence.

The supertaxon encompassing Old Orkhon, Old Uyghur, Karakhanid and Oghuz-Seljuk will henceforward be called the **Southern** (super)taxon due to its original location south of the Altai and Tian Shan Mountains.

### **Notes on the confusion about y-/J- in Oghuz and Kimak**

In this sub-chapter we briefly should consider the controversy concerning the "flickering" pronunciation of the Turkic word-initial *J-/y-*, which becomes particularly unstable when it comes to the Kimak-Kypchak-Tatar subtaxon. [We should remind again that */J-/* herein transcribes a consonant approximately similar to the English *.j*]

As we have mentioned in the very beginning, Proto-Kimak partly lost its original Proto-Great-Steppe word-initial *\*J-*, which began to mutate into *\*y-*, although this transition has never been conclusive throughout the Kimak languages. For instance, the original *\*J-* survives in Karachay-Balkar; whereas in Kazan Tatar it was preserved before- *i-* (hence Kazan Tatar *Jir* "earth", *Jil* "wind"), but changed to *y-* before other vowels (hence Kazan Tatar *yafraq* "leaf", *yul* "road", *yilan* "snake", *yörek* "heart"). Moreover, *\*J-* survives in North Crimean Tatar and Ural Tatar before any vowels.

The allophonic variation between *J-* and *y-* are also reported in East Bashkir [source: proficient speakers (2011)], and many other Kimak-Kypchak-Tatar languages.

Besides that, Mahmud al-Kashgari claimed that there existed a *y-* : *J-* or ' [zero or an Arabic hamza] correspondence both in Oghuz and Kypchak.

*For example, the Turks [=the Karakhanid Turks] call a traveler yalkin, whereas they [Oghuz and Qifchaq] call him 'alkin. The Turks call warm water yilig suw, whereas they say ilig with the 'alif. Likewise, the Turks call a pearl yinchu, whereas they call it Jinchu. The Turks call the long hair of a camel yigdu, whereas they call it Jugdu. [Diwanu l-Lugat al-Turk(c. 1073)]*

*The Uguz and Kifzhak say the words beginning with y- as J-: ul mani Jatti(he reached me) instead of yatti. At-turk say suvda yundum (I bathed in water), whereas they [Oghuz and Qifchaq] say Jundum. Amongst the Turks and the Turkman, there exists this constant rule. [Diwanu l-Lugat al-Turk (c. 1073)]*

Despite this quote, al-Kashgari also confusingly cites a good dozen of Oghuz words beginning with the *y-*, as if, either what he had said earlier no longer applied to them, or the reader was supposed to make the *y-to-J* substitution for himself. The latter seems likely, taken that this substitution was recommended by al-Kashgari in the beginning of his book.

On the other hand, it is unclear why /*J-*/ is mostly absent from the modern Oghuz-Seljuk languages including Standard Turkmen. However, at a closer look, we find that /*J-*/ does exist in many dialects of Turkmen, specifically, Karakalpak Turkmen, and as the /*J-*/ > /*d'-*/, /*t'-*/ mutation in Saryk, Yomud, Ersar Turkmen [see *Sravnitel'naya gramatika tyurkskikh yazykov. Fonetika* (1984) p. 261 ], which makes al-Kashgari claims more substantiated.

Hence, we have the Old Russian *zhenchug'* (first attested c. 1160) and Hungarian/*JönJi*/ "pearl", originally from Chinese, but in fact borrowed either from Cuman-Polovtsian that must belong to the Kimak subtaxon or from Bulgaric, though the latter option is much less likely.

### **Conclusions:**

It seems that the /*J-*/ and the /*y-*/ were interchangeably used both in the early Oghuz and Kimak languages. Both subgroups still retain the wobbly allophonic usage, which may vary across different dialects. The real life pronunciation, which sometimes differs from a textbook version or a written literary standard, adds some credibility to Mahmud al-Kashgari's account from the 1070's.

## **The Orkhon-Karakhanid subtaxon**

### **Orkhon-Karakhanid as a valid subtaxon**

The Orkhon-Karakhanid subtaxon is thought to include, among the most significant representatives, Orkhon Old Turkic, Old Uyghur (Kara-Khoja), and Karakhanid. The relatedness of Khalaj to this group is less evident (see a separate discussion of Khalaj below).

Note that in some sources, such as Lars Johanson's *Turkic Languages*, Starostin's Starling database, Orkhon-Yenisei Old Turkic, Old Uyghur (Kara-Khoja) and Karakhanid are all confusingly viewed as one and the same language. We should stress that, in theory, there might be no direct connection between them (or even between Orkhon and Yenisei Old Turkic inscriptions), and it actually stands to be demonstrated that they all belong to the same subtaxon.

### **Orkhon-Karakhanid history and geography**

All the languages of this subtaxon were located to the south of a relatively narrow passage that separates the Tian Shan ridges from the Altai-Sayan mountain system. Therefore, these languages belong to the desert and semi-desert habitat of Dzungaria, Tarim Basin, Mongolian Gobi and southern Mongolia.

As we mentioned above, the *Kul Tegin*, *Bilge Kagan* and other Orkhon inscriptions describe the *Tür(ü)ks* (the speakers of Orkhon (Old Turkic)) as enemies of the Kyrgyz, Tatars and many other local ethnicities (circa 550 AD), so we may expect a physical and linguistic separation of Orkhon Old Turkic from other Turkic branches by the time, when the events described in these inscriptions were taking place. This predicts that the Orkhon-Karakhanid languages must have appeared at least five-to-eight centuries before that date, judging by the minimum reasonable amount of glottochronological time required for a language formation, and taken that the *Tür(ü)ks* should have spoken a dialect at least slightly different from their adversaries.

### **Orkhon-Karakhanid phonology**

The following presumably innovative mutations are known in the Orkhon-Karakhanid phonology:

(1) A distinct and stable \*S- > y- innovative mutation:

cf. Chuvash *s'ichê*, Sakha *sette*, but Orkhon Old Turkic *yeti*, Karakhanid *yeti* "seven"; or Sakha *süreq*, Tuvan *chüreq*, but Orkhon Old Turkic, Karakhanid *yüreq* "heart".

This process left few traces of the original \*S- in any of the Orkhon Turkic descendants and is clearly attested as /y-/ in Karakhanid by Makhmud al-Kashagri;

(2) The presence of an intervocalic -G- and the final -G:

cf. Chuvash *pôver*, Sakha *bïar*, Kypchak *\*bawur*, *bawir*, but Orkhon Old Turkic and Karakhanid *baGir* "liver";

Orkhon Old Turkic *taG*, Karakhanid *taG*; Uzbek, Uyghur *taG* (from Karakhanid), as well as in Oghuz-Seljuk: Turkish, Azeri, Turkmen *daG*; Khakas, Tuvan, Tofa *taG* (an independent formation), but Proto-Kimak *\*taw*, Kyrgyz *too*. It is rather hard to tell whether the is an archaism or innovation, but judging by the coincidental usage of -Gin the Altay-Sayan subgrouping, it may be an archaism.

**Note:** the loss of -G-, -G in western Turkish and Gagauz as in the modern Turkish *olaJaGim* > *oliJa:m* "I will be" is a historically recent and completely different phenomenon.

(3) The retention of **the intervocalic sonants -n-, -ng-, -m-**, where the Great Steppe and Altay-Sayan have -y- or zero.

Cf. Karakhanid *süNük*, Orkhon Old Turkic *süñök* and Turkmen *süñk*, Azeri *sümük*, but Proto-Kimak *\*süyek* "bone", Tuvan, Khakas, Kyrgyz *sööök*. That this is an archaic retention is evident at least from Sakha *unuox* and Chuvash *shâmâ*, where the sonants are also retained.

(4) The **retention of the intervocalic -D-** as in Orkhon Old Turkic and Karakhanid *DaDak* "foot", *uDî* "sleep", which was possibly pronounced as an alveolar /ð/ as opposed to the languages of the Great Steppe which all have a /-y-/. That this is an archaism is evident from Khakas *azax*, Chuvash *ura*; the lenition process finally led to its loss in the Central supertaxon.

(5) Possibly, the lack of sonorization in -k-, as in Old Orkhon Turkic *säkiz*, *toquz*; Karakhanid *säqiz*, *toqu:z*, Proto-Oghuz-Seljuk *\*sekiz*, *\*doquz*, but Proto-Kimak *\*segiz* "eight", *\*toGuz* "nine", and Kyrgyz *segiz*, *toGuz* with a voiced consonant;

(6) Possibly, the retention of the word-final -b/-v as in Orkhon Old Turkic *sub*, Old Uyghur *suv*, Karakhanid *suv*; Turkmen *suv*; (also Kimak-Kypchak-Tatar *\*suw*), but Sakha *u:*, Tuvan, Tofa *suG*, Khakas *suG*, Altay *su:*, Kyrgyz-Kazakh *su:*; Oghuz-Seljuks *su*;



(7) Possibly, the  $-S^* > -ch$  word-final transition, where the original palatalized  $*S$  was stabilized through fortition:

cf. Chuvash *vis's'ê* Sakha *üs, kü:s*, Tofa *üish, küsh*, Tuvan *küsh*, Khakas *üs, küs*, but Orkhon Old Turkic *üch* "three", *küch* "force";

Chuvash *ê's'*, Sakha *is-*, Tuvan *izh-*, Khakas *is/iz-*, but Proto-Orkhon-Oghuz-Karakhanid (Turkic, Azeri, Turkmen, Uyghur, Uzbek) and Proto-Great-Steppe *ich*- "drink".

### Orkhon-Karakhanid grammar

The following features are notable in grammar:

(1) The retention of a consonant in the **verbal copula** *er-* / *är-* as opposed to *e-* / *i-* in Oghuz-Seljuk, Kimak-Kypchak-Tatar, Sakha, Altay-Sayan, etc. Cf. Old Uyghur *ärür*, Orkhon Old Turkic *er-*, and Karakhanid *ol* (a pronoun that might have substituted the original copula). It is also retained in Yugur (see below))

(2) The retention of **the instrumental case with the ending** *-(n)in*, *-(n)in*. Albeit substituted by *-la* in Kalaj. It is also present in Sakha (*-nan*), Khakas (*-naN*, *-neN*), therefore it is probably archaic;

(3) The formation of **the directive case ending in** *-Garu*, *-gärü*, found in Orkhon Old Turkic, Old Uyghur, Karakhanid; although absent from Khalaj;

(4) The use of ***-Gai*, *-gey*, *-qay*, *-kêy* as the Future Tense** in Old Uyghur, Karakhanid, Khalaj, and Chagatai (where it apparently comes from Karakhanid). This suffix is also found in a rather disjointed fashion in Yugur, Cuman-Polovtsian, Tofalar, where it might have emerged from the Optative Mood independently.

### Orkhon-Karakhanid vocabulary

The lexicostatistical research of Orkhon Old Turkic, Old Uyghur and Karakhanid is absent, except for the results provided by Anna Dybo for Swadesh-110 (2006), which attempt to position Old Turkic somewhere at the bottom of the Great Steppe subtree, which is probably due to the abundance of archaisms. As already stated elsewhere, a 100-word list would be just insufficient to differentiate between finer points in a classification, so its use in controversial cases with a small mutual separation seems unacceptable.

Judging by a notable lexical differentiation of the Oghuz branch, we should infer that Orkhon-Karakhanid must have been at least just as differentiated, therefore we cannot exclude the possibility that Orkhon Old Turkic and Karakhanid were quite different languages.

### **Conclusions:**

Based on (1) the unavoidable geographic separation by the Sayan-Altay and Tian-Shan mountain system; (2) some exclusive features in phonology and grammar not shared by either "Siberian" or Great-Steppe languages; (3) and some arguable evidence from a brief lexicostatistical study, we may infer that Orkhon-Oghuz-Karakhanid, or *Southern*, was a separate branch in its own right similar to the Altay-Sayan or Great-Steppe languages. The inference is mostly based on the exclusion of other subgroups, rather than on positive factual evidence, because the direct documentation, such as the full-fledged Swadesh lists or accurate pronunciation guides of Old Turkic, are absent due to the extinction of languages in this subgroup.

### **Khalaj is probably an offshoot of South Karakhanid**

Apparently, no other question in formal Turkology has been filled with so many nonsensical overestimations as the position of Khalaj that was considerably exaggerated in the studies of Gerhard Doerfer. Nevertheless, there is truth to some of those claims: being the only present-day survivor of the extinct Orkhon-Karakhanid branch, Khalaj stands conspicuously distinct against the background of the local Seljuk and Iranian languages.

In the present research, Khalaj is viewed as *an offshoot of the southern dialect of Karakhanid or Old Uyghur* with considerable and predictable Azeri and Persian posterior influence.

The first clear and concise account of Khalaj was made by Minorsky [V. Minorsky, *The Turkish dialect of Khalaj*, Bulletin of the School of Oriental Studies, London (1940) ] during his stay in central Iran in 1906. Minorsky's views on Khalaj classification were quite reasonable and rather contained.

However, according to [Gerhard Doerfer](#), who revisited the Khalaj speakers in 1968-73 and then published a series of articles in 1974-78, Khalaj is some kind of a fundamental Turkic language, similar in this respect to Chuvash or Sakha. This idea has been spreading like a Turkological virus, apparently because Khalaj is so remote that no one knows anything about it and no one has been able to revise that judgment with most information on this language coming only from Minorsky's and Doerfer's articles. [Note that Doerfer also denied the existence of the Altaic family.]

As Oleg Mudrak noted in his morphostatistical study of Turkic languages (2009), Doerfer's position on the subject "*rather reflected the joy of discovering a language retaining the archaic -d-*", than an outcome of an objective and unbiased analysis.

In any case, based upon the early studies by Minorksy, we must conclude that certain peculiarities of Khalaj do set it aside from other nearby languages.

On one hand, the presence of the following grammatical and phonological features mark Khalaj as a typical Seljuk language similar to Azeri:

- (1) the *-ior-* present tense marker, presumably from Azeri;
- (2) the 1st person plural *-ik* marker, e.g. *-d-ik* in past tense, presumably from Azeri;
- (3) the typical Seljuk *b- > v- > 0* mutation (as in *\*bar > "var"*, *\*bol > "uol"*), evidently as in Azeri and Ottoman Turkish.
- (4) the use of *da:l* for negation instead of *\*e(r)mes*, which is a typical Oghuz-Kimak feature (see above); presumably borrowed from Azeri.

On the other hand, *Khalaj does seem to exhibit some archaic features, not found in the Oghuz-Seljuk languages but typical of Orkhon-Karakhanid*, such as:

- (1) the **unvoiced word-initial t-, k-**, as in *ta:G* "mountain", *ki:echä* "night", *kez, kiz* < *\*köz* "eye";
- (2) the retention of **the intervocalic -D-** as in *hada:q* "foot"
- (3) the retention of **the word-final -G** in disyllabic words, as in *ha:chuG* "bitter", *sa:ruG* "orange";
- (4) the retention of the **-YmYz verb marker**, which is completely atypical of the Great Steppe languages, but typical outside of them, for instance in Orkhon-Karakhanid;
- (5) the striking retention of **the ür- copula** "to be" as in *ärtil* (as opposed to the Turkish and Azeri *idi*), apparently as in Karakhanid, Old Uyghur and Old Turkic, as well as in Yugur and Salar;
- (6) the full retention of **-qa, -ga in the dative case**, which is not typical of Seljuk-Oghuz;
- (7) the **locative case with the -cha / -che** ending, rarely found in Seljuk-Oghuz;

(8) the **future tense with *-(i)Ga***, which is normally found in Orkhon-Karakhanid (*-Gai, -gei, -qai, -kei*, etc), though it also developed, apparently independently, at least in Tofalar and Cuman-Polovtsian.

As you can see, most of these features are grammar-related, which provides significant backup for the hypothesis that Khalaj is not an Oghuz-Seljuk language and was originally related to a different stem.

At the same time *nearly all of these features are consistent with the Karakhanid origins of Khalaj*. Particularly, the future tense with *-Gay-* and the *är-* copula seem to point exclusively to Orkhon-Old-Uyghur-Karakhanid and no other branch of Turkic languages.

As to the the lexical perspective, a lexicostatistical study performed by A. Dybo (2006) viewed Khalaj as being distantly related to Turkish and Azeri, which marks it as belonging to the Southern supertaxon.

Subjectively speaking, Khalaj words are usually recognizable and the Khalaj texts are more or less readable using the knowledge of Turkish or Azeri, which is evident from the very fact that Minorsky, the earliest researcher of the language, was able to pick up a great deal of words and expressions in his first field study. If Khalaj constituted a separate branch similar to Sakha, the glottochronological differentiation would be so strong, that the language would become completely incomprehensible without special preparation.

However, Doerfer took several steps further insisting on a unique position of Khalaj among any other Turkic languages.

Based on his research, the following features are usually cited as the evidence for the uniqueness of Khalaj:

- (1) the retention of presumably primary long vowels, as in Turkmen;
- (2) the above-mentioned retention of the intervocalic *-D-* as in *hada:q* "foot";
- (3) the above-mentioned usage of the conjugated copula *är-*;
- (4) the frequent usage of the case ending in *-cha* in different meanings, including in the meaning of the locative case, as it is presumably found in Old Turkic;
- (5) the occasional persistent presence of the mysterious *h-* before vowels;

Nevertheless the presence of these traits in Khalaj can be explained in a number of ways:

- (1) The long vowels may turn out to be a recent development, considering that the language vocalism tends to change rather fast and often varies across different dialects. Neither do we have any significant evidence confirming that the long vowels must have necessarily been part of

Proto-Turkic. On the other hand, they might have been part of the Southern supertaxon, whose vocalism is poorly studied due to the deficiencies of the Arabic or Orkhon-Yenisei writing system. The latter explanation seems to be more likely, considering that we know that the long vowels are also present in Turkmen, thus presumably constituting a quite normal Oghuz feature, which may go back to Orkhon-Oghuz-Karakhanid.

(2) The retention of the intervocalic *-d-* may easily be explained by reminding that *Karakhanid* was also preserving the intervocalic *-D-* as in *aDaq* until about the 13th century, therefore this feature is also explainable from Karakhanid.

(3) The retention of the archaic *är-* copula "to be, is" is a very interesting phenomenon, which is by no means exclusive to Khalaj, as we do find it at least in Karakhanid, early Chagatai, Old Uyghur, Orkhon Old Turkic, Yugur, and Salar. Cf. Khalaj *Konduru-chä är-t-im* "I was in Kondurud", *koy-är* "it is black", *yol-ï (yol-u?) pis är-ti* "the road was bad / muddy", *var-m-or-um-är* "I'm not going" (note the archaic usage of the verb *var-* in the meaning "to go, leave" is no longer common in modern Turkish and Azeri). As already noted above, this feature too seems to identify Khalaj as part of the Karakhanid subtaxon.

(4) Additionally, both Minorsky and Doerfer found the usage of *-cha* in Khalaj in the locative meaning, as in *u-cha* "in the sleep", *yan-ï-cha* "on its side". On this basis, Doerfer (1971) assumed that this was the ending of an ancient locative case. However, there seems to be no locative case with *-cha* in Old Turkic, only a comparative case with *-cha* in Orkhon Old Turkic and Old Uyghur. Therefore the locative case in Khalaj may be an independent development based upon the usage of the comparative *-cha / -che* when answering the how-question, e.g. "how? where? — in the sleep". It has the same common adverbial meaning as, say, in modern Turkish *gün-ler-je* "during these days", *chojuk-cha* "in a childish way", etc. However, this point appears to be somewhat inconclusive, and we must admit that the usage of *cha- / che-* in the locative might indeed represent a sort of unique trait, though there are no objective reasons to believe it goes back to Proto-Turkic.

(5) As to the famous word-initial *h-* problem, despite all the suggestions that it might be remnant of a Proto-Turkic feature, a careful comparison with other Altaic languages reveals that this notion does not hold water.

The Mongolic and Tungusic-Manchu languages have extremely complex rules for the word initial *x-/ h-/ 0-* correspondences (sometimes known as the Ramstedt-Pelliot law). An */h-/* may be present in one language but then disappear in another, or mutate into an */f-/*. As a matter of fact, there's no conclusive proof that the Middle Mongolic */h-/* can be traced back to a */\*p-/*. Quite to the contrary, in many cases it seems to correspond to the Turkic */k-/* or */q-/*, e.g. Middle Mongolian *hula'an*, Khalkha *uLa:n /ush'an/*, Dongxiang *xulan*, Dagur *xula:n*, Bonan *fulaN* "red", cf. Chuvash *xerle*, Turkic *qizil* < *\*qiRil* (also see [The Mongolic / Tungusic Language Cluster](#) herein). The Tungusic word *\*xalgan* "foot" (as in Evenk, Negidal) is apparently akin to the Middle Mongolian *kol* "foot", probably having nothing to do with the *\*adaq*. On the other hand, Orok *palzhan* "foot" might in fact be a secondary development from *xalgan* > *falgan* > *palzhan*, whereas the Nanay *begdi* may be a different word

altogether, akin to the Proto-Turkic *\*but*. As one can realize, that is all very complicated and far from obvious. So it is very unlikely that anyone could have shown that the Khalaj *h-* is in regular correspondence with any of the Altaic roots.

In the *Etymological Dictionary of Altaic Languages* by Starostin, Mudrak, Dybo (2003), the authors seem to have arrived at the same conclusion:

*"One may note that this prothetic h- is very frequent before long vowels and before the following -j- and -v-. However, the rules are not strict, and in general the emergence of h- in Khalaj is unpredictable. The absence of h- in Khalaj is therefore an almost certain sign of \*0- in Proto-Altaic, so its presence there may be either original or secondary. We shall thus continue to use Proto-Turkic forms without the initial \*h- "*

Furthermore, the hypothesis of *h-* being a unique survivor retained exclusively in Khalaj is simply not statistically viable. If Khalaj were so archaic, other languages would also exhibit similar traces of the Proto-Turkic *\*h-*.

Consequently we arrive at *the hypothesis of the prothetic origin of word-initial /h-/ in Khalaj*, which will find a quite plausible corroboration below.

In fact, the obvious explanation can be found in the very beginning of Mahmud al-Kashgari's work (1073), which includes the following passage:

*"The people of Khutan [= the city of Khotan along the southern ridge of the Taklamakan desert that still exists] and Kanjak (Känchäk) [= another city further to the east] substitute the 'alifs [= the word-initial hamza plus the letter "A"] by an h (ha:). That is why we do not consider them among the Turks [=pure Karakhanid Turks], for they introduce something foreign into the Turkic speech. For example, the Turks call the father 'ata, whereas they say hata, the mother — 'ana, whereas they say hana." [Diwanu l-Lugat at-Turk].*

Surprisingly or not, this observation was made as early as the original Minorsky's article (1940) with its first description of Khalaj, so the whole thing must have been evident right from the start but then overrun by Doerfer's assumptions.

We can see quite explicitly from this passage that *Khalaj* might initially have been an offshoot of a South Karakhanid dialect spoken near *Khotan*, but then it may have traveled west along the Silk Road until it finally settled in Persia, where it survived the Mongol invasions which contributed to the disappearance of the original Khotan dialect of the Karakhanid Khanate.

Therefore, the word-initial *h-* in *Khalaj* is evidently a prothesis, but how possibly was it produced?

At first glance, the development of an *h-* may possibly be explained by the presence of an Arabic substratum in South Karakhanid, since the vowels in Arabic are preceded by a hamza that may have finally developed into an /h/. The presence of the Arabic substratum in Persia and the Tarim Basin should hardly be surprising, considering this was the Golden Age of Islam and the period of the Middle Caliphate, when Arabic was ubiquitous and could have reached Khotan via the Silk Road.

However, there seems to be no other specific evidence of exclusive Arabic influence in *Khalaj*. The fact that a different language could have been spoken in Khotan is corroborated by Marco Polo (1275) who mentions that there were several languages spoken along the southern part of the Tarim Basin. And as a matter of fact, we do know the names and even have a detailed linguistic description of some of these languages: evidently, these were **Khotanese and Tumshuqese**, belonging to the Saka subgroup of the Iranian languages.

Khotanese (or *Khotanosakan* in the Russophone literature) is rather well-attested and well-studied by Iranologists, and indeed we do find **the prothetic /h/ in Khotanese** at least some cases, cf. the following examples:

- (1) Khotanese *handara*: versus Avestan *antarê* "other";
- (2) Khotanese *hu:dva* versus Avestan *uba-* "both";
- (3) Khotanese *häysä* versus Avestan *iza-* "leather, skin";
- (4) Khotanese *halstä* versus Avestan *arshti-* "lance, javelin";

Evidently, the word-initial /h-/ in *Khalaj* finally finds explanation from the Khotanese materials.

Moreover, the word-initial /h-/ is also present in some of the Azerbaijani dialects, where its origin is rather unclear and may be a secondary formation connected with the *Khalaj* substratum.

### **Conclusion:**

We must conclude that *Khalaj must have formed along the southern edge of the Taklamakan desert on the basis of the local dialects of Karakhanid or Old Uyghur. The presence of the word-initial /h-/ can be easily explained from the Khotanese substratum which was characterized by a prothetic formation of /h-/ before vowels.*

From the southern towns of the Taklamakan desert, Khalaj could have subsequently traveled towards Persia by moving along the Silk Road thus preserving the southern Karakhanid dialect for posterity. In Persia, it came into contact with the Seljuk languages and the Persian superstratum.

Khalaj *cannot* constitute an early diversified branch of the Turkic languages, as Doerfer suggested, though it still has a few unique peculiarities lost in other branches. The Orkhon-Karakhanid hypothesis of the Khalaj origin still makes it a rather archaic language occupying a stand-alone position as compared to other Turkic languages (outside Turkish and Azeri) mostly due to an early separation of the Southern supertaxon before the 2nd century BC.

### **The Yugur-Salar subtaxon**

#### **Yugur seems to be ancient**

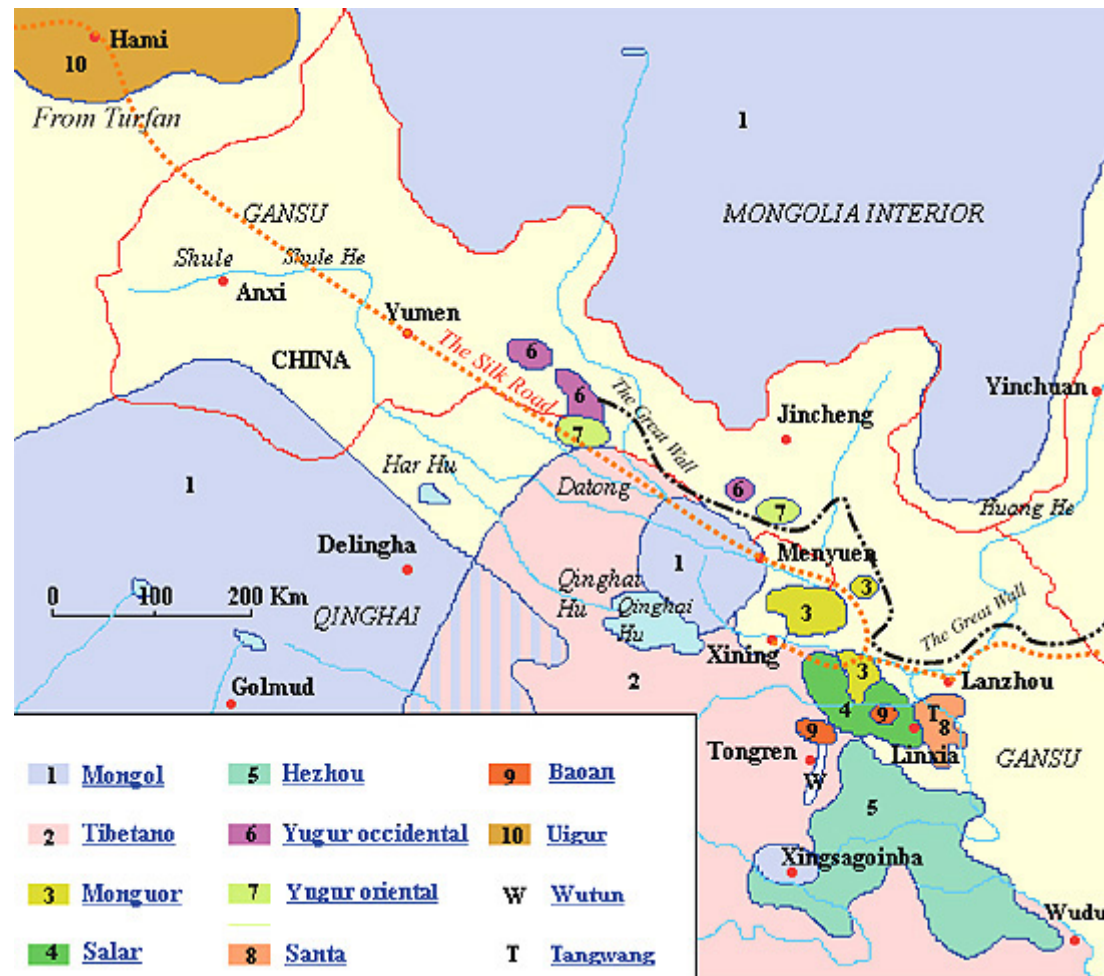
In the present study, the Yugur and Salar languages are regarded as part of *a strongly creolized early Turkic branch*, probably distantly related to the Orkhon-Karakhanid subtaxon, with some intense posterior influence from the nearby Chinese, Mongolic and Tibetan languages.

#### **Yugur history and geography**

Yugur and Salar were originally located on the outskirts of ancient China, in the vicinity of the Silk Road protected by the Great Wall in the north and the Qilian Mountains in the south. From the historical and geographical perspective, they look like an outcome of merchant settlements along the Silk Road where it enters China.

Note that part of the Yugurs were finally Mongolized and thus formed a small separate Mongolic ethnic group known as *East Yugurs* or *Shira Yugurs* speaking a Mongolic language of the same name, which is sufficient to conclude that the Mongolic influence in the region must have been very strong.





An ethnographic map of Yugur and Salar [proel.org (2010) (Only a few features added.)]

Speaking of the origin of Yugur, several simple conjectures could be made.

First, we could suggest that the Yugur people could possibly be emigrants to Turfan and Ganzhou from the Orkhon Valley civilization, known as Eastern Uyghur Kaganate, that was said to be destroyed in 840 AD by the Yenisei Kyrgyz tribes, therefore, in theory, Yugur might be related directly to Orkhon Old Turkic. However there exist certain geographic difficulties in migrating from the the Orkhon Valley to Ganzhou, which is about a 600-800 miles away and separated by the Gobi Desert.

Secondly, according to Tenishev [E. Tenishev, B. Todayeva, *Yazyk zhyoltykh ujghurov (The language of the Yellow Uyghurs)*, Moscow (1966)], the legends of Yugur people claim that part of their tribes moved about 500 miles from Turfan to Ganzhou after the introduction of Islam, which would have resulted in a geographically natural migration along the Silk Road from the Kara-Khoja Khanate (where Old Uyghur was supposed to be spoken). This second hypothesis likewise explains the origin of the ethnonym Yugur / Uyghur and it is also more geographically viable.

As a third option, we might assume that the Yugurs may have emerged from the intermingling with the Yenisei Kyrgyz population that must have lived north of that area, near Lake Zaysan, and thus consequently Yugur might be related to Proto-Altai-Khakas or Proto-Great-Steppe languages. Note that they still had to travel an enormous distance from Zaisan to Ganzhou, covering about 1000 miles through the Dzungarian Desert.

Finally, a fourth suggestion would be that Yugur is a completely independent and poorly-classified branch of the Turkic languages.

### **Yugur phonology**

The Yugur phonology is often terribly modified in contrast to other Turkic languages suggesting strong Chinese influence having accumulated over many centuries.

Just like many other languages in the region, Yugur developed the semivoiced / aspirated consonantism, so the European voiced-unvoiced letters no longer reflect pronunciation, whereas the reading of consonants is rather similar to the pinyin orthography.

A notable and quite unique feature of Yugur is **the formation of -sh- after /i/** as in *isht* "dog", *ishki* "two", *bisht* "louse". A similar phenomenon is also found in Uyghur and its dialects, and seems to be a regional innovation absent in other branches.

However despite these striking mutations, most phonological traits in Yugur are either *typically Proto-Turkic* or *typically Proto-Southern*, pointing towards Orkhon-Karakhanid:

(1) The **\*S to y- mutation** is a typical feature of Orkhon Turkic and Karakhanid, as in Yugur *yuldis* "star" as opposed Khakas *\*chiltis*, Altai *d'ildis*, Kyrgyz *Jildiz* (though the Kimak-Kypchak tribes also developed a partial \*S > \*y mutation, as described above).

**Note:** On the other hand, some examples from Tenishev *The Language of the Yellow Uighurs* (1966) show that a word-initial Mandarin-type /tsh'-/ affricate may also be present in some of the Yugur dialects in this position, but this is hardly confirmed in other sources.

(2) The presence of **an intervocalic nasal -N-** as in Orkhon and Karakhanid, e.g.

Yugur *simik*, Chuvash *s'ômô*, Old Orkhon Turkic or Karakhanid *süNök*, Uyghur *söNek*, Azeri *sümük*, Turkmen *süNk*, but Kyrgyz *sö:k*, Kazakh *süyeq*, Uzbek *suyoq*, Tatarsöyaq "bone"; this seems to be a Bulgaro-Turkic archaism, whereas the /-m-/ from the nasal /-N-/ may be a later development;

Aslo cf. Yugur *moNis*, Old Orkhon Turkic and Karakhanid *müNüz*, Uzbek *mugiz*, Uyghur *müNgüz*, but Tuvan *müyis*, Khakas *mü:s*, Standard Altay *mü:s*, Proto-Kimak-Kypchak and Kazakh-Kyrgyz *\*müyüz* "horn";

(3) The presence of **an intervocalic -G-** as in Proto-Orkhon and Karakhanid and their descendants, e.g. Yugur *paGir*, Old Orkhon Turkic *baGir*, as opposed to *Khakas pa:r*, Altai *buur*, Kyrgyz *boor*, Proto-Kimak *bawir* "liver";

Similarly, the retention of **the word-final -G** as in Yugur *taG*, *quruG*, Old Orkhon Turkic and Karakhanid *taG* "mountain", *quruG* "dry", but Altai *tu:*, *gurgak*, Kyrgyz *to:*, *gurGak*, Proto-Kimak *\*quru*. Though, this feature does not exclude the Khakas *taG*, *quruG*;

(4) The retention of **-lq-, -rq-**, e.g. Yugur *kurgak*, Old Orkhon Turkic *qulqaq*, but Khakas *xulax*, Tuvan *kulak*, Kyrgyz *kulak* "ear", etc;

(5) The retention of **the intervocalic -\*D- > -z-** as in *azaq* "foot", *Guzuruq* "tail", cf. Karakhanid *ađak*, *quđruk*, Old Orkhon Turkic *aDak*, and Khakas *azax*, *quzurux*, Tuvan *quduruq*. The purely superficial coincidence with Khakas might have led earlier researchers to believe that Yugur may be connected with the Yenisei Kyrgyz languages. However, this transition is not necessarily bears any relation to Proto-Khakas, where a similar -\*D- > -z- transition is rather unique and not shared in Tuvan. Rather it seems to be just a natural lenitional mutation that could have occurred independently, and thus per se cannot demonstrate the relatedness between Yugur and Proto-Khakas or the Altay-Sayan languages;

On the other hand, Yugur is characterized by a rather heavy consonantism with the retention of *-d-* and *-t-* where the light *-l-* is supposed to be found in the Southern branch representatives, which reminds of Altay, Kyrgyz and other Altay-Sayan-related languages, and either implies a posterior influence or a retention from the Proto-Turkic level.

### **Yugur grammar**

The Yugur grammar is largely simplified and often phonologically unrecognizable. It looks strongly creolized, a far cry from the generally familiar, typical Turkic grammars, cf. such instances as Turkic *men bar-ma-dim* vs. Yugur *men par-ma-ti*; Turkic *sen bar-ma-din* vs. Yugur *sen*

*par-min-ti* "I / you did not go" or Turkic *balam* vs. Yugur *mlaN*"my child" or Turkic *men yaz-Gan-man* vs. Yugur *men tiz-Gak er* "I am writing" or the uniquely Yugur *men tut-qish-tro* "I will catch (it)".

The strong phonological and grammatical changes in Yugur compared to other Turkic languages sort of remind of French in contrast to other Romance languages, but Yugur mutations sometimes seem to be even more pronounced. We should keep in mind active contacts with Chinese and Dongxiang (=Santa), however many modifications in morphology can hardly be explained by external influence.

Nevertheless, there are the following interesting retentions which seem to be indicative of the Yugur relatedness to Old Orkhon Turkic or Old Uyghur, cf.:

(1) the *i:re, yer* copula;

(2) the use of a **Future Tense with the -Gu** marker (instead of the Optative Mood as in "Siberian" or Kyrgyz);

(3) a peculiar presence of the Future Tense with *-qir* (in Yugur, Salar) and *-qish*(Yugur), which is probably akin to the Old Turkic construction *ROOT + qil/qish-* "to make do smt" (causative aspect).

On the other hand, some of the typical Orkhon-Oghuz-Karakhanid features seem to be absent, e.g.:

(1) The *-Gan-* suffix is used instead of the Southern *-mish-*, the latter being virtually unknown.

(2) The 2nd pers. plural *seler* "you", which is typical of Altay-Sayan, Kyrgyz, Uyghur, cf. Khakas *sirer*, Uyghur *siler*, but not in the Southern supertaxon;

Furthermore, consider the following table:

Tense	Yugur	Old Orkhon	Old Uyghur	Karakhanid	Khakas
Future Tense	-Gu, -gu, -Go, -go; -Gī, -ge, -kī, -ke	-tachī, -dachī; Giy (rarely)	-Gay, -gey	-Gay, -gey, -qay, -kêy	Gai/gei, qai/kei = Optative Mood

Perfect Tense	-Gan (usually Narrative Past); the -miş participle or tense seem to be entirely unknown	-miş-, -miş; -Gan-	-miş-, -miş-;	-miş-, -miş; -Gan-, -gen-, -qan, -ken-	-Gan/gen, -xan/ken
plural	-lar, -nar, -dar, -tar	-lar	-lar	-lar	-lar, -nar, -tar
you	seler	siz	siz	siz	sirer
copula	i:re	er-	ärür	ol (3rd pers. copula)	–

Moreover, there are some peculiar grammatical features that also seem to extend beyond the Proto-Southern level:

(1) **The -taG comparative case**, e.g. *mīn-taG* "like me", apparently very archaic, since the comparative case survives only in the Yakutic and Kimak branch, cf. Sakha *-ta:Gar*, Kazan Tatar *-day, -tay*.

(2) Yugur seems to be **one of the very few Turkic language outside Chuvash that retain *ku* "this / that; he / she / it"** mostly used as a personal pronoun "he, she". It is also found as *kini* in Sakha. The odd *ku* pronoun is evidently an Altaic retention, also well-known in Korean and Japanese. However, Yugur also has the usual Proto-Turkic *pu* "this" (absent from Chuvash).

### Yugur lexis

Certain common isoglosses in Yugur are shared with Orkhon Old Turkic and Karakhanid, but most of them seem to be even earlier archaisms going back to the Proto-Turkic level, e.g.

(1) Yugur *bezik*, Orkhon Old Turkic *beDük*, but Khakas *uluG*, Altai *d'a:n*, Kyrgyz *ulu:*, *choN* "large, great"; the former is a Bulgaro-Turkic archaism;

(2) Yugur *iliG*, Old Turkic *elig*, but Khakas *xol*, Kyrgyz *qol* "hand"; the former is a Bulgaro-Turkic archaism;

(3) Yugur *emiG*, Old Turkic *emig*, Tuvan *emig*, Sakha *emiy*, but Khakas *im-Jäk*, Kyrgyz *em-chek* "breast" with a diminutive suffix; the former is an archaism;

(4) Yugur *uzi*, Old Turkic *uDī*, Khakas *uzi-rGa*, Sakha *utuy*, but Kazan Tatar *yoklarGa*, Kyrgyz *ukto:*, Uyghur *uxli-mak* "to sleep"; the former is an archaism, perhaps Bulgaro-Turkic or Altaic, cf. Mongolian *unta-*;

(5) Yugur *yaGmīr*, Old Turkic *yaGmur*, Altai *jaNmīr*, but Kazan Tatar *yaNGīr*, Kyrgyz *jamgīr* "rain"; the former is a Bulgaro-Turkic archaism;

(6) Yugur *yaG*, Turkmen *ya:G*, Uyghur *yaG*, Karakhanid *yaG*, but Kyrgyz *may*, Kazan Tatar *may* "fat"; the former is a Bulgaro-Turkic archaism;

(7) Yugur *yildis*, Uyghur *yiltiz*, Uzbek *ildiz*, Sakha *silis*, but Turkmen *kök*, Great-Steppe *\*tamīr* "root"; the former is an Altaic archaism, cf. Middle Mongolian *ündü-sün*;

Nevertheless, the glottochronological study by Anna Dybo (2006) positioned Yugur into the Khakas-Altai subgrouping, as if it were related to the Yenisei Kyrgyz tribes. For this reason, below we will try to find words pointing specifically to northern languages, such as the Great-Steppe Sprachbund or Altay-Sayan, and show that they contain no exclusive shared innovations.

(1) Yugur *yu*, Altai *üy*, Kyrgyz *üy*, but Orkhon Old Turkic, Karakhanid *ev*, Khakas *ib* "home, house"; actually, this Yugur word may turn out to be an independent formation produced in the following way: *\*iv > \*yiw > yu*, taken that the prothetic word-initial *y-* is a common Yugur feature, and there is no direct phonological correspondence with the Great-Steppe *\*üy*.

(2) Yugur *yīrla-*, Kyrgyz *jīrlau*, Kazan Tatar *jīrlarGa*, Bashkir *yīrlau*, Nogai *yīrlaw*, but Altai *kozhoNdor*, Khakas *īrlirga*, Tuvan *īrla:r*, Sakha *illa:* "to sing"; one may initially think that this is a Kimak borrowing, but just like in the example above the word-initial *y-* is a secondary formation in Yugur that bears no relation to the Kimak languages, hence Proto-Turkic *īrla-* > Yugur *yīrla* "to sing", so the resemblance must be coincidental.

(3) Yugur *qīl-*, Uyghur *qil-mak*, Kyrgyz *qīlu*, Bashkir *qīliu* "to do"; even though this word is most typical in Kyrgyz-Chagatai languages it can also exist outside of it, and it seems to be a Proto-Turkic archaism, judging from the Tuvan *kīlir* "to do";

(4) Yugur *törtun*, Altai *törtön*, Sakha *tüört uon*, Tuvan *t.örtēn* "forty", but *\*qīrq* in any other Bulgaro-Turkic, e.g. Karakhanid *qīrq*, Kyrgyz *qīrq*, etc. However this must be an independent regular formation in Yugur that has nothing to do with the "Siberian" taxon. We may suppose that at some point Yugur seems to have lost all of its decade numbers and had to rebuild them from scratch; this is corroborated by the innovative formation of *īshk-on* "20" and especially *ūch-on* "30" which do not exist anywhere outside Yugur. However, note that the familiar *yīGīrmo* "20" is also present in Yugur, perhaps constituting a later borrowing;

(5) Yugur *kazdıq*, Sakha *qatiriq*, Khakas *xastirix* "(tree) bark"; the presence in Sakha shows this must be an archaism;

### **Conclusion:**

The geographical position of Yugur along the eastern end of the Silk Road and along the Chinese boarder sheds some light on its remarkable origins. Judging by the great variety of Mongolic and Tibetan languages in the area and the presence of peculiar features in the Yugur grammar and vocabulary, Yugur must have formed from *alinguistic intermingling of many Silk Road travelers during the late Middle Ages*. In other words, Middle Yugur can probably be regarded as *a type of a creolized language that emerged as a result of the interaction among an unknown Proto-Turkic substratum, the Old Uyghur of Kocho, the local Tibetan and Mongolic adstrata and the Mandarin superstratum*.

We found *no specific innovations relating Yugur to Altay-Sayan or the languages of Great-Steppe. Most phonological, morphological and lexical features of Yugur seem to be very archaic and pointing either to the Proto-Southern or Proto-Turkic level*.

In any case , at this point, we were unable to trace the Yugur origins with sufficient precision. Nevertheless, the collected information is sufficient to view Yugur as *a rather independent taxon within Turkic Proper*.

### **Salar has little to do with Oghuz, but quite a lot to do with Yugur and Uyghur**

#### **Salar history**

According to legends, Salar seems to be an eastern Chagatai migration branch that originated either from the Uyghur cities of the Taklamakan Desert or even the Samarqand city in Uzbekistan. The Salar people arrived in China most likely by moving along the Silk Road after the dissolution of the Karakhanid Khanate during the Mongol rule of the 13th-14th century. Their legendary date of arrival is circa 1370 which matches the rise of Tamerlane in Uzbekistan.

#### **Salar cannot be related to Oghuz-Seljuk directly**

Being a remote and forlorn language located far and deep in Central Eurasia, Salar, just like Khalaj and Yugur, has been surrounded by a number of traditional misconceptions. A common widespread belief unsupported by much reasonable evidence is that Salar is an Oghuz language.

Not all scholars accepted this view, however, and there has always existed certain controversy about this issue.

Nicholas Poppe, for instance, in the *Remarks on the Salar language* (1953) analyzed its vocabulary and phonology using Potanin's field materials, and came to the conclusion that Salar must be an "East Turki dialect", probably meaning that it must be part of the Chagatai-Uyghur language-dialect continuum. (He ignored, however, the striking differences in Salar, which should make it almost completely unintelligible to any other Turkic speakers).

Tenishev, who studied Salar in vivo in 1957, ambiguously supported its traditional classification as Oghuz despite the many facts to the contrary that he himself had provided [E. Tenishev, *Stroj salarskogo jazyka (The structure of the Salar language)*, Moscow, (1976)].

A classification of Salar within the Chagatai subtaxon has been suggested (at least) by Karl Menges in *The Turkic Languages and Peoples* p. 60. (1962, published in 1968).

On the other hand, Arienne Dwyer argued for the more traditional "Oghuz" positioning of Salar in her article [Arienne M. Dwyer, *Salar: A Study in Inner Asian Language Contact Processes, Part I: Phonology*; // *Turcologica*, herausgegeben von Lars Johanson, Band 37,1 (2007)].

The following features in Salar are often cited as typically Oghuz:

(1) The western dialects of Salar exhibit the ***b* > *v* Seljuk-type transition** (as in Salar*vu* "that, s/he"; Salar, Turkish, Azeri *var*). Yet, that cannot be viewed as an intrinsic and specific Oghuz feature, neither is it actually Oghuz (only Seljuk), and can easily be seen as a parallel phonological development.

(2) The presence of the archaic *-miş-* audative past tense (?), though the *-Gan-drand* the *-Gan-var* tense still seem to be more common. However, this feature is not uniquely Oghuz, it can also be found in Old Uyghur, Karakhanid, Chagatai and is essentially an archaic retention from the Southern supertaxon (see above).

(3) The presence of several Oghuz words, such as *el* "hand", *saG* "right", *beyle* "thus", *se:chi* "sparrow" [all mentioned by Reinhard Hahn in *The Turkic Languages*, edited by Lars Johanson, Eva Csato]. However, *el* seems to be also found in Chagatay (uncertain) or may rather be an independent formation from *eli*, the latter being known in many local languages, cf. Yughur *lG*, Karakhanid *elig*, Uyghur *ilik* (dialectal), Old Uyghur *elig*. The *saG* "right" from "healthy" is connected to the purity of the right hand in Islam and may have developed independently or found its way from the Oghuz languages. As to *beyle*, it is also found in Karakhanid as *byle* [Borovkov, A.K. *The Lexis of the Middle Asian Tefsir of the 12-13th centuries*, Moscow (1961), quoted via the Starling database]. By the same token, *seche* "sparrow" is also found in Karakhanid [cf. *sechä* in Mahmud al-Kashagari's *Divan*].



There are also a few features in Salar that could, in theory, demonstrate some similarity to Turkmen, the most typical representative of the Oghuz subtaxon, e.g.:

(1) The lack of personal conjugation in some tenses in Turkmen (such as Turkmen *-Jag* (future), *-makchi* (intention), *-malī* (obligation), which, however, are all absent in Yugur-Salar.) Nevertheless, the loss of grammatical markers cannot be viewed as a shared innovation, and, in Salar, it is obviously a result of the secondary contact with Mandarin and Mongolic languages. Actually, a similar process of losing personal conjugation — apparently under the influence of the local languages — has also occurred in Khalkha-Mongolian and to some extent in Yugur.

(2) A peculiar usage of *-yok* to express negation in verbs in some tenses, as in Salar *ROOT + yoxtur* (Present) and Yugur *ROOT + qīsh + yoq-tīr* (Future II), distantly similar to the Turkmen *ROOT + a + personal marker + ok* construction as in *yaz-a-m-ok* (= "I haven't written", lit. "no my writing"). But evidently, this feature finds a local Yugur parallel, and its analogy in Turkmen may be purely coincidental.

Furthermore, the comparison to the typical Oghuz shared innovations demonstrates their absence in Salar and therefore shows the lack of any direct connection between Salar-Yugur and Oghuz languages (see Oghuz features above for reference):

(1) No trace of *deyil/deGil*, which is a standard form of negation in Oghuz and Kimak-Kypchak-Tatar. A more archaic *\*emes(tir)* is used instead in Salar and Yugur;

(2) The dative with the *-ga /-a* ending, which is not typical of Oghuz, where only the *-a*-ending is used almost exclusively. But cf. *-Ga, -ge, -qa, -ke* (without *-a*) in Yugur;

(3) The forms of the genitive case do not coincide with those in Oghuz, being similar only to those in Karachai-Balkar, the Lobnoor dialect of Uyghur, and some of the Uzbek dialects (see Tenishev (1975)), with the Uyghur and Uzbek dialects evidently being the original source of these mutations in the Tian Shan area;

(4) The system of verbal tenses is quite similar to Yugur, it lacks any personal endings, and has nearly nothing to do with Turkmen, Azeri, or Turkish, except for the most basic forms recognizable in all the Turkic languages;

(5) There is no *siz* pronoun "you" in Salar-Yugur; cf. Salar *sele(r), sile(r)* for plural (as in Kyrgyz, Uyghur) and *sen* for polite reference being used instead; note that the personal pronouns of the 1st and 2nd person are rarely borrowed or substituted.

(6) There is a notable lack of any typical Oghuz lexical innovations, such as Oghuz-Seljuk *\*kök* "root", cf. Salar *sachax*; Oghuz-Seljuk *\*choq* "many", cf. Salar *köp*; or any typical Oghuz phonological innovations, such as Oghuz-Seljuk *\*boynuz* "horn", cf. Salar *moNiz*.

(7) The audative past tense with *-miş-* does exist, but the *-miş-* marker does *not* seem to join adjectives or nouns, which seems to be a distinguishing feature of the Seljuk-Oghuz languages (and Uzbek dialects, where it is apparently from Karakhanid).

(8) The *ROOT + por/par/padır = Present Tense* grammeme bears no relation to the Oghuz Present Continuous with *-yor-*, as Tenishev claimed, but is apparently akin to the present tense *ROOT-ıp-par* in Yugur, where *par(dır)* is akin to Karakhanid *\*bar* "be present". Hence, evidently, *ROOT + yox-tur* in the negation of verbs in Salar.

(9) There is no "Oghuz voicing" in Salar, as many researchers thought, since most word-initials are either unvoiced or *semi-voiced*, which is sometimes incorrectly reflected in writing as fully voiced consonants by European linguists. A simple explanation of this phenomenon is that the Salar phonology tends to follow the Mandarin system: *strong aspirated vs. weak semivoiced*. The degree of voicing may vary creating the impression of full voicing (noted by Tenishev).

This is the usual areal feature common to many languages of the Far East (Yughur, Tuvan, Mongolian, Korean, etc), not necessarily because of the direct contact with Mandarin but rather due to the longstanding mutual interaction of local languages and the formation of a common linguistic area, especially as far as the phonology is concerned. Furthermore, Tenishev says in his own words:

*The system of the Salar consonantism is so drastically different from the South Turkic (Oghuz) system, which was supposed to exist for the Salar language in the past, that one involuntary arrives at a conclusion of its secondary, posterior origin, and its dependence upon the neighboring languages, such as Chinese, Dongxiang, Tibetan. [E. Tenishev, *Stroj salarskogo jazyka (The structure of the Salar language)*, Moscow, (1976)]*

Consequently, Tenishev explains how the phonological systems of Mandarin and Dongxiang (=Santa) could have affected the Salar languages.

He does not go as far as rejecting the "Oghuz hypothesis", however, probably unwilling to go against the mainstream view of his time, but many of the facts he explicitly mentioned do point in that direction.

### **Salar cannot be an offshoot of the Great-Steppe languages either**

By the same token, it was shown in *A List of Phonologically Dissimilar Basic Words in Central Asian Turkic Languages* (above), that Salar can hardly be directly related to other Great Steppe subtaxa, at least because of the following discrepancies:

- (1) the presence of the –G-, -G velar as in Salar *paGiz*, *taG*, cf. Kimak-Kypchak-Tatar *bawir* "liver", *tau* "mountain", Kyrgyz *boor*, *too*;
  - (2) Salar *emes* "not", cf. Kimak-Kypchak-Tatar *tügel*;
  - (3) Salar *uxla-* "to sleep", cf. Kimak-Kypchak-Tatar *\*yukla*, cf. Kyrgyz *uktoo* (no match in either case);
  - (4) Salar *yi-*, Kimak-Kypchak-Tatar *asha-* "to eat", cf. Kyrgyz *Je*;
- etc.

### **Salar may share some features with Uyghur**

The only apparent proximity to any other familiar Turkic language can be found in Modern and Old Uyghur, Chagatai and Karakhanid.

(1) Just like Yugur, Salar exhibits the Karakhanid-Chagatai *y-*, which has been shown herein (see *Introduction*) to be a late Southern innovation.

(2) Both Yugur and Salar share a number of peculiar developments, such as an additional sound after /i/ or /i/ in "two": Yugur *shigi*, *ishke*, *ishqi*: Salar *ishki*, *ichki*. Curiously, this development also frequently appears in spoken Uyghur dialects, but never in writing [reported by a proficient speaker].

Nevertheless, we cannot position Salar in the same subtaxon with Karakhanid because of the absence of certain typical Karakhanid archaisms in Salar. Cf. the following examples:

- (1) Karakhanid *ev* "house" : Salar *oy*;
- (2) Karakhanid *uDa-* "to sleep" : Salar *uxla-*;
- (3) Karakhanid *yapurGaq* "leaf" : Salar *yarfix*, etc.

Moreover, we know from historical sources, that Salar must have emerged in the 14th century already after the disappearance of Karakhanid. That leaves us with Uzbek-Uyghur-Chagatai as the only possible source of phonological influence, with the eastern Uyghur dialects being the likeliest candidates for Salar's closest linguistic neighbors. Cf. the following examples:

- (1) Uyghur *öy* "house" : Salar *oy*

- (2) Uyghur, Uzbek *uxla-* "to sleep" : Salar *uxla-*;
- (3) Uyghur *müNgüz* "horn": Salar *moNiz*, as opposed to Uzbek *mugiz*, *shoz* (from Kypchak and Persian respectively);
- (4) Uyghur *süNäk* "bone" : Salar *senix*, as opposed to Uzbek *suyak* from Kimak-Kypchak-Tatar;
- (5) Uyghur *beGir* "liver" : Salar *paGir*, as opposed to Uzbek *zhigar*;
- (6) Uyghur *qo:saq* "belly" : Salar *xusax*, as opposed to Uzbek *qorin*.

Therefore, Salar seems to reflect some Chagatai or Uyghur phonological and lexical influence.

### Evident similarities in the Yugur and Salar grammar

As noted above, *the main influence in Salar in fact comes from Yugur*, and as Tenishev briefly asserted [idem], "*The very same order of tenses is observed in Yugur*". Indeed, the similarities in the verbal systems in both languages are striking; some of them are listed in the table below.

Tense	Yugur	Salar	Comment
Present Progressive	<i>ROOT+ip+par</i>	<i>ROOT+por</i>	This tense is rather innovative, probably from <i>*par/var</i> "there is", as it follows from the examples in the other Salar tense <i>ROOT + Gan var</i> as well as from <i>par-dr</i> "there is"; the relatedness to the verb <i>*bar-</i> "to go" has also been suggested, though Tenishev for some reason assumed that <i>-par</i> is from the Oghuz <i>-yor-</i> .
Aorist	<i>ROOT+ar</i> (Future)	<i>ROOT+ir/er</i> (Present-Future)	Common to all Turkic (no taxonomic value)
The "Yugur" Future	<i>ROOT+qir</i>	<i>ROOT+qur</i>	Apparently, a unique Yugur-Salar innovation
The Simple Past	<i>ROOT+te</i>	<i>ROOT+Je</i>	Common to all Turkic languages, but still phonologically innovative, including the striking absence or degradation of personal endings.

The <i>Gan-</i> Past	<i>ROOT+Gan+tro</i>	<i>ROOT+Gan+dır</i>	Common outside of Oghuz-Seljuk, but the addition of <i>-dır</i> or <i>-tro</i> is rather innovative.
----------------------	---------------------	---------------------	--

The bizarre lack of personal conjugation markers in verbs in Salar and partly in Yugur can naturally be ascribed to the Sino-Tibetan or Mongolic influence.

Note: Concerning Mongolic, Tennishev notes [idem], "most Mongolic languages, including Dongxiang, lack personal conjugation. It is only present in the Kalmyk and Buryat languages, and the Bargu-Buryat and Oyrot dialects of Mongolian." This observation may work as a further corroboration for the existence of some sort of a typological Sprachbund near Mongolia and northern China.

Also, cf. the apparently exclusive matches in indefinite pronouns Yugur *qim-er*, Salar *kem-ter* "someone", Yugur *nier*, Salar *naN-tir* "something".

Both Salar and Yugur use the *ira(r)* copula akin to the Old Uyghur *ärür*, which is used after nouns and adjective much in the same way as the English is. This is a quite peculiar feature, especially considering a similar phonological development from /ä-/to /i-/ in both Yugur and Salar. A similar usage has also been found in Khalaj (see above). The presence of *-r-* in this root can be regarded as a typical Orkhon-Karakhanid archaism.

	Yugur	Salar	Comment
<b>copulas</b>	<i>er, ere, ire</i>	<i>ira, irar;</i> <i>iter, itir, ider; ideroN</i> (except the 1st person); <i>tir, dir, tir, dir;</i> <i>shi, shê</i> < Mandarin	Cf. Old Uyghur <i>ärür</i> , Khalaj <i>är</i> ; According to Tennishev, the Salar <i>itir</i> = <i>ira</i> + <i>tir</i> (a double copula), just as in <i>emes-tir</i> , <i>emes-er</i> (a negative copula)
<b>examples</b>	<i>xo p'er k'i:se i:re</i> "[we] all one people are"	<i>wu pirinige oy iter</i> "this our house is"; <i>men xon iter</i> "I the-khan am"; <i>inJi avu ira vu</i>	Also, used in Salar much in the same way as "right, it is" in English. <i>Man ka'cha yanshaGani idero? — Ider!</i> "What I said, is it right? — It is." <i>Men pichtigeni ira mu? — Ira.</i>

		"a young(man) still he is"; <i>putaGi pir ideroN</i> "their roots <i>one</i> are"	"What I wrote, is it right? — It is."
--	--	---	---------------------------------------

In a nutshell, the notable matches in grammar clearly demonstrate the close relatedness between Yugur and Salar.

### **Salar lexis**

There's no detailed lexicostatistical study of Salar, except the one in Anna Dybo's work, who again places Salar near Turkmen, which is highly dubious. A superficial overview of the Salar Swadesh-110 (collected by Starostin (1991)) suggests that this language contains many unusual lexical innovations and would only be poorly intelligible by the speakers of Oghuz languages.

To confirm the low level of mutual intelligibility between Salar and Seljuk languages, we will provide a link to this lovely (and well-performed) traditional Salar song with very simple lyrics: <http://www.youtube.com/...>

*usher ya, mA(nya) (maNa) ushEr-ya!*

*salar (seler) mAnya ushEr*

*yaNi pizgen zOrakh-ne tAxiner pAshime*

*akokO akokO akokO, pAshime*

*usher ya, mA(nya) ushEr-ya!*

*salar mAnya ushEr*

*Ichim tikh-ken tonimne gi:ir pONime*

*akokO akokO akokO, poN(i)me*

*usher ya, mA(nya) ushEr-ya!*

*salar mAnya ushEr*

*Apam AlGan Ishtan-nE ki:ir di:zime*

*akokO akokO akokO, ti:zeme*

*usher ya, mA(nya) ushEr-ya!*

*salar mAnya ushEr*

*Izem Etken xAim-ne gi:ir ayaqE*

*akokO akokO akokO, ayaqE*

A broken Turkish translation (with the maximum usage of cognates) would look something like this:

*Üşür ya! Bana üşür ya! (Oh look at me! Gather around me!)*

*Siz-ler bana üşür! (Etrafımda toplanın!) (You all gather around me!)*

*Yeni beze-yen şapkayı (süslenmiş bir şapkayı) taşır[ım] (giyerim) başıma (The newly ornated hat, I shall wear on my head)*

*Annem[-in] dik-en palto[sun]u (diktiği paltoyu) giyer[im] bedenime (The by-my-mother sewn coat, I shall wear on my-self (my body))*

*Babam[-ın] al-an pantolon[un]u (aldığı pantolonu) giyer[im] dizime (The by-my-father bought pants, I shall wear on my knees)*

*"öz-üm" ed-en ayakkabılar[ın]ı (kendi yaptığı ayakkabıları) giyerim ayağa (The by-my-self-made shoes, I shall wear on [my] feet)*

Despite some intelligibility, most Turkic words in the song lyrics are barely recognizable. Actually, nowhere outside Chuvash and "Siberian" do we find so many strong phonological, lexical and grammatical changes — that is, changes at all the levels of language structure — as we do in Yugur and Salar, which makes their taxonomic positions quite questionable and rather distant from most other Turkic subgroups.

### **Conclusion:**

Consequently, based on the strong grammatical evidence, we must conclude that *Salar and Yugur belong to the same subgroup*, whereas *Salar is probably based on the Yugur substratum*. Additionally, Salar retained much of the Chagatai vocabulary and phonology of the arrivals from the Tarim Basin which helped to preserve some mutual intelligibility with other languages of the Southern taxon.

Therefore, Salar seems to be a sort of ethno-linguistic seam formed on the interaction border between the language of the Yugur merchants and the newly-arrived refugees or economic migrants from the Chagatai Khanate. These new settlers may have been coming in several waves of migration, so the process of supplanting and creolizing the local Yugur substratum in Ganzhou could not have been an overnight event, probably taking several centuries.

The modern Salar is likely to be a Chagatai-Yugur creole that emerged as an admixture of the Yugur substratum, the Mandarin and Mongolic adstratum, and the Uyghur-Chagatai superstratum. As the Ganzhou kingdom Yugur speakers gradually acquired new Chagatai vocabulary and some of the new grammatical features, the early Salar rose as a distinct language with the Yugur grammatical basis but the modified Uyghur-Chagatai vocabulary and the Mandarin-Mongolic phonology.

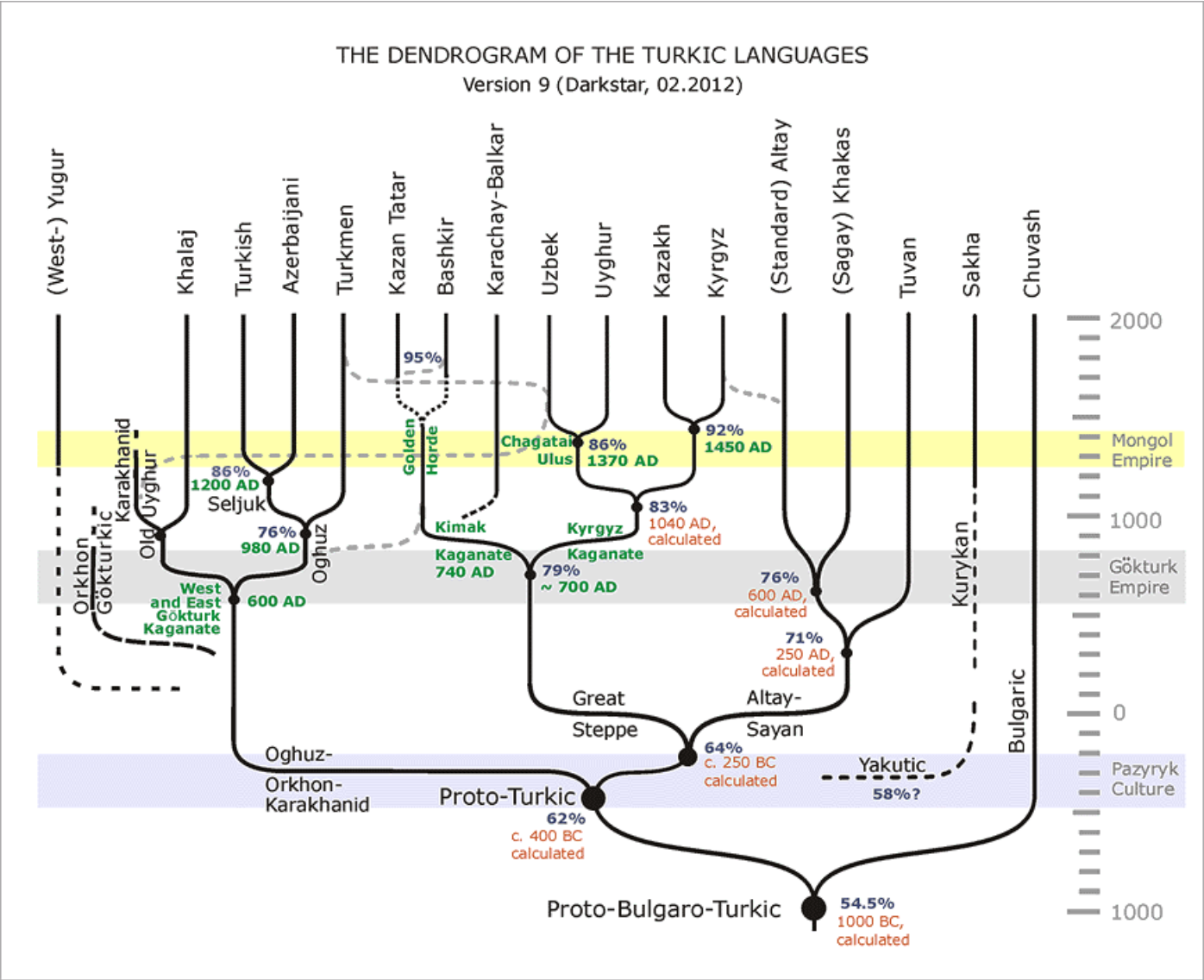
However, some questions concerning the origins of Salar and Yugur still remain, and the matter of their exact taxonomic position is far from clear.

#### **4. The Resulting Internal Classification of Bulgaro-Turkic Languages**

##### **4.1 The Genealogical Classification of Bulgaro-Turkic Languages**

As an outcome of the present research, we can now build a plausible tree of the Turkic languages including their internal mutual influence. The resulting dendrogram should look approximately as follows (only the languages included into the lexicostatistical study plus Khalaj, West Yugur, and Old Turkic are shown in this figure):





The dendrogram of the Turkic languages (2012)

## **4.2 The Taxonomic Classification of Bulgaro-Turkic Languages**

Taxonomic classifications are often regarded as being of secondary importance, since they cannot reflect all the complexities of real phylogenetic relationships, however they are still useful in many situations, for instance when classifying languages in a list. In any case, based on the kinship shown in the above dendrogram, as well as other lexical, phonological, morphological and geographical evidence provided and discussed in this publication, the Turkic languages can be subdivided into the following taxa:

### **BULGARO-TURKIC**

#### **BULGARIC**

##### **(1) VOLGA BULGARIC**

(1.1) **Chuvash** (including Chuvash and its dialects)

#### **TURKIC (PROPER)**

The sometimes accepted term "Common Turkic" is used mostly in Anglophone sources, and is best to be avoided because of its inconsistent association with such meanings as "a language common to all Turks", "commonplace, ordinary Turkic", "a common Turkic conlang", etc. Turkic in the strictest sense of the word may rather be named *Turkic Proper* or just *Turkic*, as opposed to *Bulgaro-Turkic*, which may sound slightly unusual in the beginning, but is generally self-explanatory.

##### **(1) EASTERN (or YAKUTIC)**

Despite a few features shared with the Central subtaxon, Yakutic must still be viewed as an independent branch of Turkic Proper because of multiple innovative differences. The few features shared with Altay-Sayan (and occasionally with Great-Steppe) should mostly be regarded as archaisms or a result of an older Yakutic substrate in the Altay-Sayan Turkic languages.

## **(1.1) Yakutic**

(1.1.1) **Yakutic** (including the hypothetical Kurykan (or Proto-Sakha), Modern Sakha, Dolgan)

The habitat of these languages is mostly connected with the Lena basin.

## **(2) CENTRAL (or ALTAY-SAYAN-GREAT-STEPPE)**

### **(2.1) Altay-Sayan**

Geographically, most of the ethnicities in this subgroup belong to the upper Yenisei and Ob basins.

(2.1.1) **Tuvan** (including Tuvan, Tofa (outdated: Tofalar), Todzhin, Soyot, Tsaatan)

(2.1.2) **Khakas** (including Sagai Khakas, Kacha Khakas, Fuyu Kyrgyz, Shor, Middle Chulym and other closely related dialect-languages). Note that *Khakas* seems to be an entirely artificial ethnonym created in the 1920's. The positions of Fuyu Kyrgyz, Shor and Chulym have not been considered in this study.

#### **(2.1.3) Altay (Turkic)**

Note that the historical name of the mountains is spelled irregularly as *Altai*, whereas the name of languages is usually spelled more regularly as *Altay*. The sub-classification of Altay dialects goes back to Baskakov and has not been revisited ever since.

(2.1.3.1) **North Altay (Turkic)** (including Kumandy, Kuu (Chelkan), Tuba)

(2.1.3.2) **South Altay (Turkic)** (including Standard Altay or just Altay (confusingly known as *Oirot* until the 1940's; the name *Altay-kizhi* "Altay people" is also applicable, albeit illogical), Teleut, Telengit).

### **(2.2) Great-Steppe**

This supergroup is supposed to include those languages that were migrating north of the Great Eurasian Barrier across the enormous territory of the Great Steppe including such areas as Jeti-Su, the Southern Ural, the Aral-Caspian region, the Volga, the Crimean Peninsula, the Kievan Rus and even as far as Lithuania and Poland. All of these tribes most likely originate from the basin of the upper Irtysh basin and the area of Lake Zaisan.

(2.2.1) **Tian-Shan** (or alternatively, *Kyrgyz-Kazakh-Uzbek-Uyghur* or *Kyrgyz-Kazakh-Chagatai* or just *Kyrgyz-Chagatai*, according to the typical representatives).

The exact original homeland of this subtaxon and its temporal period are unclear, but it was probably situated somewhere between the Altai and Tian-Shan Mountains. By the 7th-8th century it must have moved to the foothills of the Tian Shan Mountains, hence the suggested appellation.

(2.2.1.1) **Kyrgyz-Kazakh** (including Kyrgyz, Kazakh, Karakalpak)

Kyrgyz was apparently affected by Altay Turkic ("Oirot") during the Dzungarian invasion of the 17-18th century, hence its frequent misplacement in other classifications.

(2.2.1.2) **Chagatai** (including possibly the hypothetical Karluk (?), medieval Chagatai, modern Uzbek and Uyghur and their dialects)

The subgroup is essentially an admixture of the old Uyghur-Karakhanid substratum with the language of Great-Steppe newcomers. It formed after the Mongol invasion of the Tian Shan in the 13th century. The name "Karluk" from Baskakov's classification is best to be avoided because our knowledge of Karluks is rather limited, and their Turkic dialect has not been preserved. On the contrary, *Chagatai* was a significant and commonly-used medieval koine in Central Eurasia, therefore its name sounds much more reasonable and recognizable as a taxonomic appellation.

(2.2.2) **Kimak** (or *Kimak-Kypchak-Tatar*, according to the most famous representatives of the Kimaks).

All of the ethnicities therein are thought to be descendant from the Kimak Confederacy (Kaganate, Khanate) situated near Lake Zaysan. The Kimaks were strongly affected by the linguistic exchange with Oghuz near the Zaysan Passage in the 7th-9th centuries. The older Baskakov's name "Kipchak" is best to be avoided due to the inaccurate and confusing inclusion of Kazakh and Karakalpak, the exclusion of Nogai, etc. Moreover, the actual Kypchaks constituted only a small part of the Kimak subtaxon apparently focused near the Kievan Rus, therefore overestimating their significance at the cost of of the Kimaks, the original progenitors of the subgroup, seems to be rather unjustified.

(2.2.2.1) **Karachay-Balkar** (including Karachay-Balkar and its dialects)

A linguistically deviating subgroup in the Caucasus Mountains, still evidently of Kimak-Kypchak-Tatar origin.

(2.2.2.2) **Golden-Horde** (including Sibir Tatar, Bashkir, Kazan Tatar, Mishar Tatar, (Caspian) Nogai, Kumyk, North Crimean Tatar, Central Crimean Tatar, Crimean Karaim, Lithuanian Karaim and other closely related language-dialects)

The formation of most of these Kimak languages is clearly connected with the rise and expansion of the Golden Horde during the 13th-15th centuries. Having formed during a relatively recent period, the Golden-Horde languages still share many common features. Due to a large number of languages in this subgroup, it has been studied rather superficially in this work.

(2.2.2.3) **Baraba-Tomsk** (including Baraba and probably Tomsk Tatar)

A very special Kimak subgrouping exhibiting certain archaic features and presently almost extinct. Tomsk Tatar has not been included into this study.

**(3) SOUTHERN (or ORKHON-OGHUZ-KARAKHANID)**

This major supertaxon includes the languages that migrated *to the south of the Great Eurasian Barrier* inhabiting the system of deserts, semi-deserts and steppes in the Tarim Basin, Dzungaria, Mongolia, Gobi and northwestern China named herein as the "*Gobi Steppe*". Many of these ethnic groups formed part of (or were closely related to) the Göktürk-Uyghur Empire of the 6th-9th century CE.

**(3.1) Orkhon-Karakhanid**

This subtaxon includes various extinct descendants the Göktürk-Uyghur Empire, such Orkhon Old Turkic, Old Uyghur, Karakhanid, with Khalaj being the only living representative. The original self-appellation of the speakers in this subtaxon was often *Tür(ü)k*.

(3.1.1) **Orkhon Old Turkic** (including Orkhon Old Turkic of the Orkhon inscriptions)

Also known as just *Tür(ü)k*, or *Kök Tür(ü)k*, or *Göktürk(ic)*.

(3.1.2) **Uyghur-Karakhanid** (including Old Uyghur, (North) Karakhanid, unattested South Karakhanid, and modern Khalaj)

### **(3.2) Oghuz-Seljuk**

This subtaxon was slightly affected by the Kimak languages near the Zaysan Passage circa the 7th-8th century CE and thereafter.

(3.2.1) **Oghuz** (including Standard Turkmen and the closely related language-dialects, namely Yomud, Ersarin, Saryn, Saryq, Chovdur, Trukhmen; the hypothetical "Early Oghuz" of the Oghuz confederacies during the 8th-10th century).

Turkmen seems to be rather strongly affected by the languages of the Great Steppe.

(3.2.2) **Seljuk** (including Qashqai, Khorasani, Azeri, Old Anatolian Turkic, Ottoman Turkish, Modern Turkish, Gagauz and other closely related language-dialects of Turkey, Iran and Azerbaijan)

The Seljuk languages apparently formed from an Oghuz dialect of the [Great Seljuk Empire](#) blended with Perso-Arabic elements between the 11th and 13th centuries.

### **(3.3) Yugur-Salar**

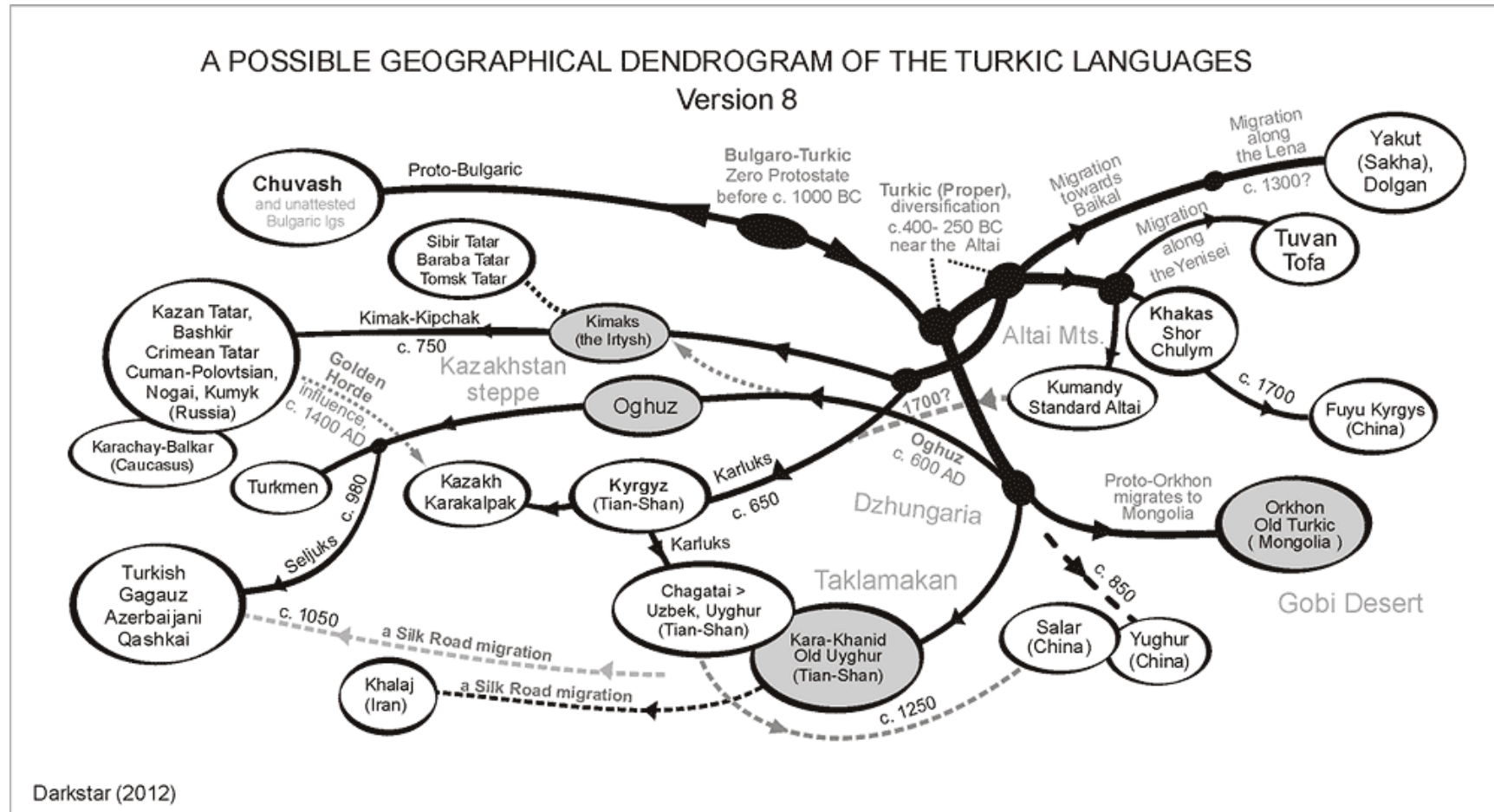
This subtaxon seems to have emerged as the result of intense intermingling of the Turkic, Mongolic, Tibetic and Chinese ethnicities near the Qilian Mountains in the Hexi Corridor where the Silk Road enters China. Despite the frequent misplacement, both Yugur and Salar seem to form a separate subgroup, most likely within the Southern taxon, though a higher and more archaic positioning may also be plausible.

(3.3.1) **Yugur** (including (West) Yugur (Yughur))

(3.3.2) **Salar** (including the West and East Salar dialects)

## **4.3 The Geographical Tree of Bulgaro-Turkic Languages**

We should also note that any attempt to build an absolutely consistent genealogical classification of closely related languages may run into considerable difficulties because of the mutual interaction among different branches and various complex wave phenomena within the tree model. For this reason, a more simple *geographical dendrogram* was additionally created that takes into consideration the migratory movement of Turkic branches. However, both dendrograms ultimately express the same taxonomic ideas.



A geographical dendrogram of the Turkic languages (2012)

For the analysis of the Proto-Bulgarian and Proto-Turkic Urheimat position see a separate article [The Proto-Turkic Urheimat & The Early Migrations of Turkic Peoples](#)

## 5. References and sources

Note that many documents, books, and articles in the list below should be available online.

### Comprehensive and standard sources

1. Lars Johanson, Eva A. Csato, *The Turkic languages*, London, New York (1998) [a standard manual of Turkic languages in English; consists of articles by specific authors]
2. *Jazyki mira: Tyurkskije jazyki (The Languages of the World: The Turkic Languages)*; editorial board: E. Tenishev, E. Potselujevskij, I. Kormushin, A. Kibrik, et al; The Russian Academy of Sciences (1996) [a detailed, authoritative edition with a brief phonological and grammatical description of each language; consists of articles by specific authors]
3. *Jazyki mira: Uralskije jazyki (The Languages of the World: The Uralic Languages)*; editorial board: V. Yartseva, Yu. Yelisejev et al; The Russian Academy of Sciences (1993)
4. *Jazyki narodov SSSR. Tyurkskije jazyki (The languages of peoples of the USSR. Turkic languages.)*; Editor-in-Chief: Baskakov, N.A.; Moscow (1966) [This is actually a thoroughly written collection of grammars and text samples of all the major languages of the ex-USSR from the "warming" period, when many outstanding works were created. Many readers have praised the quality of this book.]
5. *Starling Database, The Turkic etymology*, [starling.rinet.ru](http://starling.rinet.ru), composed by [Anna Dybo](#) [pronounced: AHN-nah de-BAW]
6. *Sravnitelno-istoricheskaja grammatika tyurkskikh jazykov. Morphologija. (The Comparative Historical Grammar of the Turkic Languages. Morphology.)*; editorial board: E. Tenishev et al, Moscow (1988) [Despite the word "grammar" in the title, this multivolume publication is essentially an



attempt at a comprehensive research of Proto-Turkic at several levels, with this particular volume dedicated to the analysis of morphology; the name is sometimes abbreviated according to the Cyrillic letters as **SIGTY**; some articles, however, seem to be too verbose and confusing for the important subjects they cover.]

7. *Sravnitelno-istoricheskaja grammatika tyurkskikh jazykov. Regionalnyje rekonstruktsii. (The Comparative Historical Grammar of the Turkic Languages. Regional reconstructions.)*; editorial board: E. Tenishev, G.V. Blagova, E. A. Grunina, A. V. Dybo, I.V. Kormushin, L.S. Levitskaja, D.N. Nasilov, O.A. Mudrak, K.M. Musajev, A.A. Chechenov, et al; Moscow (2002)

8. *Sravnitelno-istoricheskaja grammatika tyurkskikh jazykov. Leksika. (The Comparative Historical Grammar of the Turkic Languages. Lexis.)*; editorial board: E. Tenishev et al; Moscow (2002) [Many lexical examples and supposed proto-forms concerning the life of Proto-Turks.]

9. *Sravnitelno-istoricheskaja grammatika tyurkskikh jazykov. Pratyurkskij jazyk-osnova. Kartina mira pratyurkskogo etnosa po dannym jazyka. (The Comparative Grammar of the Turkic Languages. The Proto-Turkic Language. The Worldview of the Proto-Turkic Ethnic Group Based on the Linguistic Data.)*, editorial board: E. Tenishev et al., Moscow (2006) [Attempts at the mythological and semiotic analysis of the Turkic lexis from the previous volume.]

10. *Etymologicheskij slovar tyurkskikh jazykov (The Etymological Dictionary of the Turkic Languages)*, E. V. Sevortyan, Vol. 1-7, Moscow (1974-2003) [Mostly known and named herein as *Sevortyan's Dictionary*, though he died in 1978. Pronounced /seh-vor-TAHN/ as an Armenian-Azerbaijani surname. It is in fact a multivolume publication prepared by a group of authors, with the earliest volume still photocopied from a typewriter, apparently due to difficulties in reprinting diacritics; the last volumes are still being prepared for publication; despite some convoluted passages and even some discrepancies with modern dictionaries, perhaps still the most comprehensive work on the Turkic lexicon]

11. *Atlas narodov mira (The Atlas of the Peoples of the World)*, Moscow (1964) [old but good, taken that ethnographic maps generally get better with the time because of the language loss]

### **Other general sources and references**

1. Sevdal Sulejmanova, *Istorija tyurkskikh narodov (The history of the Turkic peoples)*, Baku (2009) [a laconic but fairly detailed chronology from an Azerbaijani author]

2. *Stepnyje imperii drevnej Evrazii (The Steppe Empires of Old Eurasia)*, S. G. Klyashtornyj , D.G. Savinov, Saint-Petersburgh (2005)

3. *Gosudarstvo kimakov IX-XI vv. po arabskim istochnikam (The Kimak State of the 9-11th century according to the Arab sources)*, Kumekov, B.E.; Alma-Ata (1972)

4. O.A. Mudrak, *Ob utocnhenii klassifikatsii tyurkskikh jazykov s pomosh'ju morfologicheskoy lingvostatistiki (On the clarification of the classification of Turkic languages by means of the morphological linguostatistics)// Sravnitelno-istoricheskaja grammatka tyurkskikh jazykov. Regionalnyiye rekonstruktsii*. Moscow (2002) [an abbreviated article published within the SYGTY, using a novel taxonomic approach to build the classification of Turkic languages]

5. O.A. Mudrak, *Klassifikatsija tyurkskikh jazykov i dialektov s pomosh'ju metodov glottokhronologii na osnove voprosov po morfologii i istoricheskoy fonetike (The classification of the Turkic languages and dialects based on the glottochronological methodology with a morphological and phonological questionnaire)*; Moscow (2009) [same as above, a full version in a separate book; only 100 paper copies in circulation]

6. O. A. Mudrak, *Yazyk vo vremeni. Klassifikatsija tyurkskikh jazykov. Istorija jazykov (The language in time. The classification of the Turkic Languages. The History of languages.)*(2009); [published as pdf at [www.turklib.ru](http://www.turklib.ru) and elsewhere as html, and a video; similar to the above, but made into a lecture for general public with a brief history of Turkic languages]

7. Anna Dybo, [Khronologija tyurkskikh jazykov i lingvisticheskije kontakty rannikh tyurkov](#) (*The Chronology of the Turkic Languages and the Linguistic Contacts of the Early Turks*) (2006?)

8. Anna Dybo, *Lingvisticheskije kontakty rannikh tyurkov. Leksicheskij fond*. (Linguistic Contacts of the Early Turks: the Lexical Fund), Moscow (2007) [the book includes a lexicostatistical analysis with a couple of dendrograms, and a detailed analysis of early borrowings into Proto-Turkic]

9. *Altajskaja problema i proiskhozhenije japonskogo jazyka (The Altaic Problem and the Origins of the Japanese Language)*, by [Sergey Starostin](#); Moscow (1991) [a dissertation that includes excellent, detailed 100-word Swadesh lists of all the Altaic languages with just a few occasional errors]

10. M. Dyachok, [Glottchronologija tyurkskikh jazykov](#) (*The Glottochronology of the Turkic Languages*), Materials of 2nd Scientific Conference, Novosibirsk (2001) [some preliminary materials, known mostly as a short online paper, however quite interesting]

11. Classifications of Turkic Languages by various authors (in Russian) [etho.org](http://etho.org)

Classifications of Turkic Languages by Baskakov (1969) (in Russian), [etho.org](http://etho.org)

12. Werner Froehlich, *Turkic glossary*, [www.geonames.de](http://www.geonames.de), (2001-2011) [some valuable lexical materials for various language groups including Turkic; the author states, "I created this site with the greatest possible care."]

13. [200-word Swadesh lists for Turkic languages](#) (en. wikipedia.org) [in fact, it is now superceded by the version published in this work, see a doc-file in [The Lexicostatistics and Glottochronology of the Turkic languages](#)]

14. Talat Tekin, [Türk Dilleri Ailesi](#) (*The Turkic Language Family*) // Genel Dilbilim Dergisi, Vol. 2, pp. 7-8, Ankara (1979) [on the mutual intelligibility of Turkic languages compared to Turkish]
15. A. Scherbak, *Sravnitel'naja fonetika tyurkskikh jazykov* (*The Comparative Phonology of the Turkic Languages*) (1970)
16. Yu. V. Normanskaja, [Rastitelnyj mir. Derevja i kustarniki. Geograficheskaja lokalizatsija prarodiny tyurkov po dannym floristicheskoi leksiki](#) (*The plant world. Trees and shrubs. The geographical localization of the Turkic homeland based on the floristic lexis data.*) // *Sravnitel'no-istoricheskaja grammatka tyurkskikh jazykov. Pratyurkskij jazyk-osnova. Kartina mira pratyurkskogo etnosa po dannym jazyka*. Moscow (2006) [a controversial article but interesting nonetheless]
17. N. A. Baskakov, *Vvedenije v izuchenije tyurkskikh jazykov* (*An introduction into the study of Turkic languages*, Moscow (1969) [Note that the work itself, acc. to the author, dates back to 1952 and several reprints and remakes under different names were made from this book, e.g. *Očerki istorii funktsionalnogo razvitija tyurkskikh jazykov*, Ashgabad, (1988). It should be explained that [Nicolay Baskakov](#) (1905-1995) was not just the leading Turkologist of the USSR, he was the brand of many Soviet Turkological studies, so many dictionaries of regional Turkic languages composed by different authors were printed with his name as a chief editor.]
18. Baskakov, N.A., *Sovremennye kypchakskije jazyki* (*The modern Kypchak languages*), Nukus (1987) [Again, mostly a reiteration of his own previous classification with particular emphasis on Kypchak, including South Altai]
19. Alexander Samoylovich, *Nekotoryje dopolnenija k klassifikatsiji turetskikh jazykov* (*Some additions to the classification of Turkish languages*, Petrograd (1922); reprinted in the collection of his works (2005)
20. Alexander Samoylovich, *K voprosu o klassifikatsiji turetskikh jazykov* (*Towards the question of the classification of Turkish languages*, the Bulletin of the 1st Turkological Congress of the Soviet Union (1926); reprinted in the collection of his works (2005)
21. *Aus Sibirien. Lose Blätter aus meinem Tagebuche* (*From Siberia: Torn pages from my diary*), Wilhelm Radloff, Leipzig, 1893 [A wonderful ethnographic description of Altay, Khakas, Kazakh, Kyrgyz, Uyghur and Uzbek people, early archaeological evidence, etc. An absolutely awesome book first hand. There exists an abbreviated Russian translation from as late as 1989.]
22. *Brockhaus and Efron Encyclopedic Dictionary*, Saint Petersburg (1906)
23. [The long and wonderful voyage of Frier John de Plano Carpini](#), by Frier John de Plano Carpini (1245-46)
24. *Forschungsreise durch Sibirien 1720-1727*, by [Daniel Messerschmidt](#) (1721-1725)
25. Mahmud al-Kashgari, *Compendium of the Turkic Dialects* (c. 1073); [an English translation (1982) by Robert Dankoff and James Kelly]
26. *The Secret History of the Mongols* (1240), translation by F. W. Cleaves (1982) [a translation from the Mongolian original]

### **Specific Turkic languages**

*Russko-chuvashskij slovar*, by M. Skvortsov, A. Skvortsova; Cheboksary (2002) (doc)

*Nutshell Chuvash*, by Andras Rona-Tas, Szeged (Hungary) (2009?)

*Etymologicheskij slovar chuvashskega jazyka (The etymological Dictionary of Chuvash)*, by M. Fedotov; volume 1-2, Cheboksary (1996) [quite helpful and enlightening]

*Chuvashskij jazyk i jego odnoszenije k mongolskomu i tyurkskim jazykam (Chuvash and its relatedness to Mongolian and the Turkic languages)*, Nicholas Poppe (1924) ([downloadable](#))

Russian-Yakut, Yakut-Russian online dictionary (22.000, 35.000 words), [www.sakhatyla.ru](http://www.sakhatyla.ru)

Brigitte Pakendorf, *Contact in the Prehistory of the Sakha, Linguistic and Genetic Perspective*(2007)

Shirokobokova, N.N. *Otnoshenije jakutskog jazyka k tyurkskim jazykam Juzhnoj Sibiri* (The relatedness of the Yakut language to the Turkic languages of South Siberia), Novosibirsk (2005) [this is essentially, a small monograph on the linguistic origins of Sakha]

*Grammatika tuvinskogo jazyka*, F. Iskhakov, A. Pal'mbakh, Moscow (1961) [a remarkably detailed grammar of Tuvan with comparative examples from other languages]

*Slovar tofalarsko-russkij, russko-tofalarskij*, V.I. Rassadin, Saint-Petersburg (2005)

*Sojotsko-buryatsko-russkij slovar*, V.I. Rassadin, Ulan-Ude (2003)

V.I. Rassadin, [O probemakh vozrozhdenija i sokhraneniya nekotorykh tyurkskikh narodov Juzhnoj Sibiri \(na primere tofalarskogo i sojotskogo\)](#) (2006)

*Orys-Khakas Slovar*; D. Chankov, Editor in Chief; Moscow (1961)

*Khakassko-russkij slovar*, composed by [N. Baskakov](#), A. Inkizhekova-Grekul (1953)

*Khakasskij jazyk*, by N. Baskakov, A. Inkizhekova-Grekul, Moscow (1953)

*Dialekty khakasskogo jazyka*, Editor in Chief: D. Patachakova, Abakan (1973)

*Russko-khakasskij slovar dla khakasskikh nachalnych shkol*, Ts. Nominakhanov, Abakan (1948)

[Series of articles concerning the origins of the ethnonym "Khakas"](#), by S. Yakhontov, V. Butanayev, S. Klyashtornyj // *Ethnograficheskije obozrenije* (1992) (in Russian)

*Fu-yü Kırgızcası ve akrabaları*, Mehmet Ölmez; Mersin (1998)

*Fu-yü Kırgızcası ve akrabaları*, Mehmet Ölmez; Istanbul (2001)

*Russko-Oyrotskij Razgovornik*, composed by V. Antonov-Saratovskiy, translated by I. Kalanakov, Leningrad (1931)

*Russko-Altajskij Elektronnyij Slovar*, by U. Tekenova, S. Tekenov, E. Tatin, (TRANS.exe) (2006?)

*Russko-Altajskij Slovar*, Editor-in-Chief: Baskakov, N.A.; Director: Kuchigasheva, N.A.; Moscow (1964)

*Oyrotsko-russkij slovar*, composed by [N. Baskakov](#), Toskhakova (1947)

*Dialekt Kumandintsev /Kumandy-Kizhi/, Grammaticheskij ocherk, teksty i slovar*, by N. Baskakov, Moscow (1972)

*Кыргызча-орусча сөздөк, Орусча- кыргызча сөздөк*, by K Yudakhin

*Grammatika kyrgyzskogo jazyka, kratkij spravochnik*, Bishkek (2002)

*Grammatika kazakhskogo jazyka v tablitsakh i skhemakh*, by L. Kulikovskaja , E. Musayeva; Almaty (2006)

*Kazakhskij jazyk*, by K. Musayev; Moscow (2008)

*Kratkaja grammatika kazak-kirgizskogo jazyka*, composed by P. Melioranskij, Sankt-Peterburg (1894) [an old Kazakh textbook from the 19th century, quite interesting]

*Russko-karakalpakskij slovar*, Editor-in-Chief: N. Baskakov, composed by Sh. Karimkhodzajev, K. Kdyrbajev, et al., Moscow (1967)

*Къарачай-Малкъар Орус-Сӕзлюк*, edited by E. Tenishev, Kh. Suyunchev; Moscow (1989)

*Obschchije svedenija o karachajevu-balkarskom jazyke (General notes about the Karachay language)*, by Ali Dzharashtiyev (2009?) [online only]

*Shkolnyj russko-kabardinskij slovar*, by Kh. Dzhaurdzhij, Kh. Syk'un; Nalchik (1991)

*Russko-tatarskij razgovornik*, composed by E. Lazareva, Moscow (2004)

*Russko-tatarskij slovar slovosochetaniy (A Russian-Tatar dictionary of word combinations)*, composed by Khanif Agishev, Kazan (1996) [a good Tatar dictionary for beginner's with many examples for each word — a world of useful info]

*Tatarcha-Ruscha Uku-Ukitu Süzlege*, composed by F.A Ganiyev, I.A. Abdulin, R.G. Gataulina, F.Ye. Yusupov; Moscow (1992)

<http://www.xatasiz.com> [A good online Russian-Tatar, Tatar-Russian dictionary with an audio database]

*Dialektnyje osobennosti yazyka sredneural'skikh tatar (The dialectal characteristics of the Middle Ural Tatars)*, dissertation by Sarmanajeva D.M.; Kazan (1950)

*Govory sibirskikh tatar yuga tymenskoj oblasti (The dialects of the Sibir Tatars of South Tyumen Oblast)*, Alishina, Kh. Ch.; avtoreferat dissertatsii [a thesis summary]; Kazan (1992)

*Dialekty zapadnosibirskikh tatar (The dialects of West Siberian Tatars)*, Akhatov G. Kh.; avtoreferat dissertatsii [a thesis summary]; Moscow (1964)

*Russko-kumyjskij slovar*, Editor: Z. Bammatov, Moscow (1960)

*Russko-nogajskij razgovornik*, composed by I. Kapayev, K. Kumratova, Stavropol (2007)

*Grammatika nogajskogo jazyka. Fonetika i morfologija (The grammar of the Nogai language. Phonetics and morphology.)*, editor-in-Chief: Baskakov, N.A.; Authors: Kalmykova, S.A., Sartseva M.F., Cherkessk (1973)

*Nogayskij yazyk i yego dialekty (The Nogay language and its dialects)*, Baskakov, N.A., Moscow (1940)

*Yazyk barabinskikh tatar (materialy i issledovaniya) (The language of the Baraba Tatars (materials and studies))*, L.V. Dmitriyeva; Leningrad (1981)  
[This is one of the very few detailed field studies of the Baraba Tatars in the 20th century, conducted in the 1950-60's. It includes legends and stories recorded from illiterate participants, grammar notes and a brief vocabulary.]

*Russko-bashkirskij slovar*, composed by Z.G. Uraksin, Ufa (2005)

*Grammatika bashkirskoho jazyka dla izuchayuschikh jazyk kak gosudarstvennyj (The grammar of Bashkir for state students)*, Usmanova, M.G.; Ufa (2006)

[\*Elbrusoid Russian-Karachay-Balkar Dictionary \(Version 2.0\)\*](#)

*Uzbekskij jazyk dlya vzroslykh (samouchitel)*, I. Kissen, Sh. Rakhmatulayev, Tashkent (1990)

*Russko-uzbekskij slovar*, Editor-in-Chief M. Ch. Koshchanov; Vol 1-2, Tashkent (1983)

*Uighur - Russian Dictionary* (an electronic dictionary for ABBYY Lingvo) (2008)

*Uygursko-russkij slovar*, Editors-in-Chief: Sh. Kibirova, Yu. Tsunvazo, Alma-Ata (1961)

*Turkmen-English Dictionary*, by Garret, Lastowka, Muhammetmuradova, et al (1996)

*Turkmenskij jazyk*, by E.Grunina, Moscow (2005)

*Kratkij russko-turkmenskij slovar*, Editors-in-Chief: M. Khazmayev, S. Altayev, Ashgabad (1968)

*Turkmence-Rusca sözlük*, Editors-in-Chief: N.A. Baskakov, B.A. Karryyeva, M. Ya. Khamzayeva, Moscow (1968)

*Samouchitel azerbajdzhanskogo jazyka*, by T. Khudazarov, Baku (2006)

*Azerbaycanca-Rusca lüg^et*, Editor-in-Chief: M.T.Tagiyev; Vol. 1-4, Baku (2006)

*Grammatika turetskogo jazyka dla nachinajuschikh*, by Olga Sarygyoz (2007)

*Turetsko-russkij slovar*, composed by R. R. Yusipova, Editor-in-Chief: T. Ye. Rybalchenko, Moscow (2005)

*Turetsko-russkij i russko-turetskij slovar*, composed by T. Ye. Rybalchenko, Moscow (2007)

*Intensivnyj kurs turetskogo jazyka*, by Yu. Scheka, Moscow (1996)

*Grammatika jazyka tyurkskikh runicheskikh pamyatnikov, VII-XII vv.*, by A. Kononov, Leningrad (1980)

*Ocherk grammatiki drevnetyurkskogo jazyka*, by V. Kondratyev, Leningrad (1970)

*Drevnetyurkskij slovar (The Old Turkic dictionary)*, Editors: V.M Nadelyayev, D. M. Nasilov, et al., Leningrad (1969)

[Türük Bitig](#), a site dedicated to Orkhon-Yenisei inscriptions

*The Turkish dialect of Khalaj*, by V. Minorsky, Bulletin of the School of Oriental Studies, London [a field study, written circa 1906, but published in 1940]

*Yazyk zhyoltykh ujghurov (The language of the Yellow Uyghurs)*, E. Tenishev, B. Todayeva, (1966) [a field study of 1958, but too concise]

*The Western Yugur (Yellow Uyghur) Language. Grammar, Texts, Vocabulary*, Martina Roos, a dissertation, Leiden (2000) [a detailed manual based on a new field study]

*Remarks on the Salar Language*, by Nicholas Poppe, University of Washington (1950's?)

*Stroj salarskogo jazyka (The structure of the Salar language)*, by E. Tenishev, Moscow, 1976 [a field study]

*Salar: A Study in Inner Asian Language Contact Processes, Part I: Phonology* by Arienne M. Dwyer; *Turcologica*, herausgegeben von Lars Johanson, Band 37,1 Weisbaden (2007)



*Arabic Etymological Dictionary*, by Andras Rajki (2002)

**2009-2013 (c)**