

Mutual Intelligibility Among the Turkic Languages

By Robert Lindsay

Abstract: *The Turkic family of languages with all important related dialects was analyzed on the basis of mutual intelligibility: (1) To determine the extent to which various Turkic lects can understand each other. (2) To ascertain whether various Turkic lects are better characterized as full languages in the own in need of ISO codes from SIL or rather as dialects of another language. (3) To analyze the history of various Turkic lects in an attempt to write a proper history of the important lects. (4) To attempt to categorize the Turkic languages in terms of subfamilies, sub-sub families, etc. The results were: (1) Rough intelligibility figures for various Turkic lects, related lects and Turkish itself were determined. Surprisingly, it was not difficult to arrive at these rough estimates. (2) The Turkic family was expanded from Ethnologue's 41 languages to 53 languages. Splitting: a number of new languages were created from existing dialects, as these dialects were better characterized as full languages than as dialects of another tongue. Lumping: a few existing languages were eliminated and re-analyzed as dialects of another or newly created language. (3) Full and detailed histories for many Turkic lects were written up in a coherent, easy to understand way, a task sorely needed in Turkic as histories of Turkic lects are often confused, inaccurate, controversial, and incomplete. (4) A new attempt was made at categorizing the Turkic family that rejects and rewrites some of the better-known characterizations.*

Acknowledgments: This paper could not be written without the generous support and kind, wise heart of Professor Suer Eker of Bashkent University in Ankara, who is in charge of the book project where this article is published. Suer also assisted in quite a bit of the research. Of those who assisted in my research, probably no one was more important than the calm, patient, brilliant Field Linguist David Anderson of the Living Tongues Institute in Salem, Oregon. The depth

and breadth of his knowledge of Khakassian and Tuvan languages is stunning and he provided a huge amount of clarification about critical issues surrounding these languages.

Professor Tamás Csernyei of Hachettepe University in Ankara provided concise and in depth critical comments on a pre-publication version of the paper. He also helped with references and formatting. Csernyei's knowledge of Turkic as a whole is breathtaking.

Sema Demir, also a professor at Bashkent, helped with some critical comments on the same early version of the paper and helped with the Turkmen section. Sohrab Dolatkhan, a post-doctorate researcher at the Ecole Pratique des Hautes Etudes at the Sorbonne, Paris, provided some clear and concise comments on the South Azeri and Kashkay lects in Iran. Tomas Ericksonas, graduate student at the Institute for Asian and African Studies, University of Helsinki has extensive knowledge of Tuvan languages. This knowledge was essential for the writing of the Tuvan sections of this paper.

Sam Poova Esla's knowledge of Iranian Turkic helped with the South Azeri section. Ackhat Maneshev provided some valuable background on little known Siberian Tatar tongues. Professor Stavros Skopeteas provided good information about the little known Urum language.

What Is the Turkic Language Family?

Turkic is a large family usually thought to consist of about 40 languages stretching from Turkey all the way to China. Most of the languages are fairly close to each other, and it's often been said that they are all mutually intelligible, and you can go from Turkey all the way to the Yakut region of Siberia and be understood the whole way.

Why a 90% Intelligibility Requirement for Splitting Languages from Dialects?

Of course all Turkic languages are not intelligible, but as with many

myths, there is a certain amount of truth there. That is because the languages, while generally not above 90% intelligible which is my requirement to be dialects of a single tongue instead of full languages on their own, do have varying degrees of intelligibility.

So where to draw a line between a dialect and a language? I choose 90% intelligibility. 90% and above, we are dealing with dialects of a single tongue, and below 90% intelligibility, we are dealing with separate languages. Once again, this is justified by common sense.

Once mutual intelligibility of two lects reaches 80%, conversations about complex, educated and technical subjects start experiencing difficult communication. Two speakers can still communicate in a basic way, particularly in face-to-face conversation where speech can be slowed and nonverbal communication can be used. But not everyone wants to talk about the weather, and if two lects cannot discuss the complex issues of the day without problems, one can make the case that now we have two different tongues.

I chose 90% because that seems to be where *Ethnologue* cuts off a language from a dialect. You will note, especially in *Ethnologue's* Mexico section where many scientific intelligibility studies have been done, that in general, lects with intelligibility of another lect below 90% are considered to be separate languages. Also in many cases, such as Aragonese and Spanish (80% mutual intelligibility or MI), Asturian and Spanish (80% MI or possibly as low as 25% MI for hardcore Asturian or Leonese), Galician and Spanish (85% MI) that once again, we seem to be looking at 90% as a cut-off figure.

Also on the Spain page of *Ethnologue*, the editors refused to split off Valencian from Catalan (94% MI), instead claiming it is merely a dialect of Catalan as opposed to a separate language as Valencian activists claim. In fact, in debates about whether or not Valencian is a separate language or a dialect of Catalan, the overwhelming scholarly consensus it is not a separate language but instead is simply a Catalan dialect.

In support of this claim, the 94% MI figure is typically trotted out as proof of Valencian's dialectal nature.

Meankeli is split from Finnish on the basis of an 80% MI with Finnish.

In addition, Victor Mair, a well-known linguist who specializes in Chinese languages, agreed that 90% is a good cutoff point (Mair 2009).

The truth is that MI in Turkic is much less than proclaimed.

The aim of this paper is to attempt to shed light on the mysterious question of Turkic MI about which many myths have sprung up.

In addition, the paper will attempt to divide the Turkic lects into separate languages and associated dialects based on either structural difference or MI or both.

How to Determine Intelligibility?

Structural difference is another method of determining whether we are dealing with two separate languages. Once two lects get divergent enough, linguists generally think we are dealing two separate languages. This way of dividing languages up based on structural divergence is not particularly controversial.

Another way to divide languages form dialects is MI. The typical rejoinder is that such measures are always arbitrary and subjective, and hence have no meaning and should not even attempt to be quantified. However, intelligent MI measurements are not generally controversial among linguists, who tend to agree on such judgments as they are based on common sense.

For those who require more rigorous science, intelligibility tests have

been tested and designed over a period of 40 years, particularly by the Summer Institute of Linguistics (SIL). SIL has refined these measures down to a fine art, and they have a high degree of scientific reliability which has been checked in multiple ways. Others have come up with their own rather ad hoc intelligibility tests, which tend to show similar results to the better tests and again agree with speakers' basic judgments.

Although structural divergence and MI are two different measures, they fall together well. This is because at a certain degree of structural divergence, two different lects start experiencing a difficult time understanding each other. Increased structural divergence leads to difficult MI as surely as day follows night.

Wild, often angry and heated debates rage, typically on the Internet, about the MI of various lects. Most of the screaming about this issue is not being done by sober-minded linguists but instead is being done and as a general rule they tend to be hotheads, cranks and irrational minded people who generally know little to nothing about linguistics.

Just because the general public which is highly ignorant of such things behaves irrationally about this question is no reason why objective linguistic social scientists should shy away from the debate. As mentioned earlier, among calm, rational scientific linguists, the issues of language vs. dialect and MI are not particularly controversial as they tend to fall in line with people's intuitive judgments and basic common sense.

How Was Intelligibility Measured in This Study?

Intelligibility was measured via conversations with native speakers and linguists working with Turkic tongues. In addition, the relevant literature was consulted and expert opinions about whether a given lect was seen as a separate language or a dialect of something else were analyzed.

In addition, linguists' judgments about structural divergence of given lects was utilized.

One critique sometimes raised is that native speaker judgment and even linguist opinions may be inferior to actual MI studies in determining MI.

There is a problem with this in that these well-founded MI studies have simply not been done on Turkic languages. The only MI studies in Turkic have been two studies testing Turkish and North Azeri. So if we are only going to use well-founded MI studies to determine MI, this paper could not have been written. Until such actual MI studies are done, it is best to simply rely on statements made by native speakers.

One problem with saying we need more formal MI studies is that these studies are not often done because few linguists are interested in doing them. Even for determining %'s of MI, we often just ask native speakers of Lect A, and they tell us they have, for example, 65% intelligibility of Lect B. Or we ask a linguist who studies the language, and they say Lect A has, for example, ~85% or 75% intelligibility of Lect B. This is just a rough estimate by the linguist.

How is MI typically determined by linguists? In determining MI, *Ethnologue* researchers first simply ask native speakers of Language A how much they can understand of Language B.

They ask different groups, males and females of all ages. There are generally few to no deliberate false statements given; however, subjects must be interviewed individually and not in groups.

One oddity is that males of Lect A often report higher MI of Lect B than females do. This is due to bilingual learning, as Lect A males have been outside the village interacting with other males from Lect B for work purposes, whereas females stay in the Lect A village and have have no contact with Lect B speakers. In MI studies, we look for

“virgin ears” and try to rule out bilingual learning as it contaminates results.

So most MI determinations in the literature are determined either by asking native speakers or by querying linguists who study the language. Actual MI studies are not often done, but when they are, they typically find the same MI as the native speakers and linguists find, or, if they differ, the MI studies actually find lower MI than the native speakers and linguists reported, as native speakers and linguists tend to overestimate MI.

There will be few to no false positives (languages spit off when they are actually dialects of a single tongue or excessive splitting) via MI determination with any method.

If native speakers or linguists say Lect A is not intelligible with Lect B, then that is the fact, since their errors will occur in overestimation rather than underestimation. So all splits based on MI queries are justified. The only problem you will have in MI queries is false negatives.

Since queries tend to be biased upwards in terms of MI, at times, speakers of Lect A will say they have full intelligibility of Lect B when in fact, if you do a study, it may come out to ~82%. So all splits will be justified, but some splits that should be made will not be made (some full languages will be referred to as dialects of another tongue). In other words, excessive splitting is unlikely but excessive lumping is much more of an issue.

How Many Turkic Languages?

One might ask how many Turkic languages there are?

Ethnologue lists 41 different languages, one of them possibly extinct.

In this treatment, I expand *Ethnologue's* 41 Turkic languages to 53 separate languages utilizing measures of MI, structural divergence and the judgments of native speakers and linguists. At least one of the 53 languages may have gone extinct recently.

The subject of MI and the consequent division of Turkic on that basis has not seen a lot of research. Hence this is a pilot study intended to stimulate new research into this question. For many of the lects analyzed below, information is lacking or controversial in a number of areas. Where this is the case, I suggest topics in need of further study and elaboration.

Internal Classification of Turkic

In addition, I attempted some tentative work at internal classification of Turkic into families and subfamilies.

Although there have been many attempts at internal classification of Turkic, there has been no generally agreed-upon classification, although Tekin's recent attempt has been popular.

I welcome all critical comments about this article from sober-minded persons, particularly native speakers of Turkic tongues, Turcologists and other linguists.

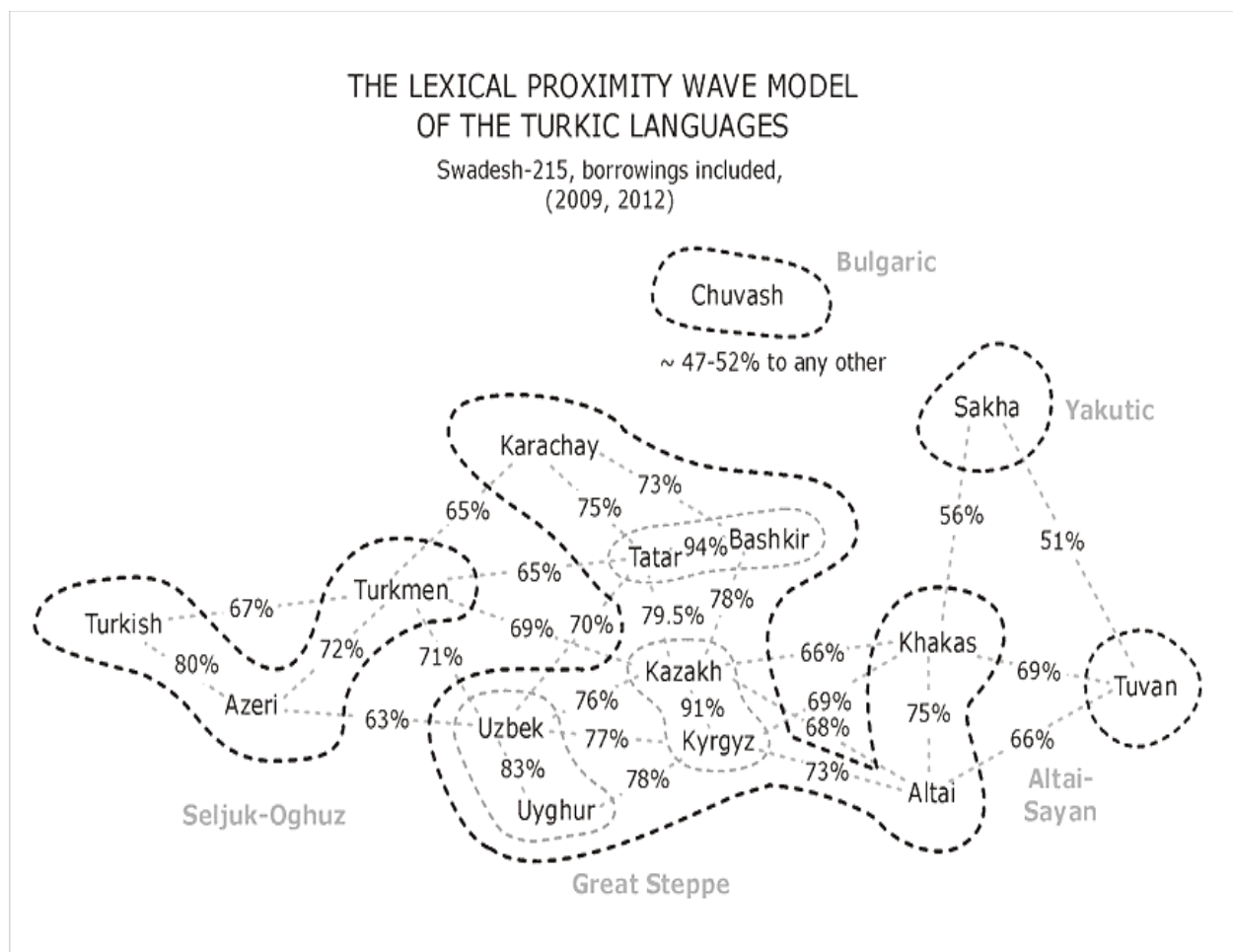


Figure 1: A Swadesh-215 list with borrowings included to represent actual speech - a good model for a rough test of intelligibility. All percentages represent cognates shared between any two lects. 90%+ cognates = dialects of a single language. 80% = ~70% intelligibility. 70% = ~40% intelligibility. 60% = ~10% intelligibility. Below 60% = essentially zero intelligibility.

Bulgaro-Turkic

Proto-Bulgaro-Turkic may have originally been situated in the East near the Proto-Turkic homeland but to the west near northern Kazakhstan, the Ishim-Tobol-Irtysh Steppe and the Baraba Steppe in

southern Siberia near northern Kazakhstan and adjacent areas in northern Kazakhstan in the Kokshetau Plateau and Kazakh Uplands around Petropavl and Pavlodar all the way south to Tengiz Koli.

To the west, it may have extended all the way over to the southern edge of the Urals near Magnitogorsk and Chelyabinsk. A possible vision of this homeland would be an area of wide, often-flooding rivers, sloughs, channels, backwaters and small lakes with ample vegetation such as swampy areas with reeds and grains and large wet meadows.

The proto-Bulgaro-Turkics may have operated small boats in this area. There were many fish and crustaceans to harvest from the waters, and the area was rich in wildlife, particularly waterfowl. Many land animals are also drawn to areas with plenty of water. The proto-Bulgaro-Turkics may well have used bows and arrows and spears to harvest wildlife.

The association of Proto-Bulgaro-Turkic with the Andronovo Culture known from archaeology and especially its Akalul and Fedorovo branches in southern to far southern Siberia is well-known. Fedorovo is known from 1100-1200 in the area from the Tobol River east through the Ishim Steppe to the Irtysh River and then east again through the Baraba Steppe to the upper Ob Basin near Tomsk. Akalul is known from the southeastern edge of the Urals to the steppes of the Ishim-Irtysh Basin from 1200-1700 BCE.

This puts Proto-Bulgaro-Turkic in the Andronovo region anywhere from 900-1300 BCE and possibly all the way from 900-1700 BCE and gives us an initial date for Proto-Bulgaro-Turkic of 3300-3700 YBP (years before present). All other things being considered, the upper theory of 3700 YBP seems best.

Some researchers have tried to associate Andronovo with Proto-Indo-Iranian. The association of Andronovo with Indo-Iranian has many problems. In particular, the Fedorovo and Akalul Cultures seem a poor

fit for Proto-Indo-Iranian which was probably in the area, particularly the southeast of the Akalul area, but they were there far earlier, from 1500-2500 BCE, and by 1500 BCE, they were already well on the way to the south towards Transoxonia. Further, I am not aware of any Proto-Indo-Iranian borrowings in Proto-Bulgaro-Turkic.

Oghuric

The time period for Oghuric's split from Proto-Bulgaro-Turkic is typically placed at 500-1000 BCE. Weighing the evidence, the best scenario is for an early split for Oghuric around 1000 BCE. They then migrated to the west, possibly fleeing some invaders of some sort.

In their new home on the Volga, the Oghurics were separated from the rest of Turkic for close to 2,500 years until contacts occurred with Tatar in the last several hundred years. This combined with a very early separation and isolation from all other early Turkic groups near the Urheimat accounts for the extreme divergence of Oghuric and its sole living remnant, Chuvash. This follows from linguistic theory whereby the most divergent member of a language family is frequently the one that is the most geographically removed from the main group.

Volga Bulgaric

[Chuvash](#) (Lewis et al 2009c), spoken by 1,325,382 people in the Republic of Chuvashia (Tishkov 2009), is the most divergent language of all Turkic. Some classifications even split Turkic into Chuvash in one high level branch and all of the rest into another huge branch, which is then further subdivided. For instance, Tekin says that all Turkic languages descend from a language called Proto-Chuvash-Turkic or pre-Turkic ([Tekin 1990](#)).

This treatment follows Tekin's model and calls the large family Bulgaro-Turkic. Chuvash has extremely poor intelligibility with all of the rest of Turkic at 0-15%. Generally it is zero. The best figure is

Uzbek-Chuvash at ~10-15%.

Macro-Turkic

Proto-Turkic

The Proto-Turkic homeland is the subject of much debate. Many scholars simply refuse to postulate on its location, stating that its location is simply uncertain. However, enough material has been gathered via historical analysis and reconstruction of Proto-Turkic that we can make a guess at the Proto-Turkic Urheimat.

A good guess would be somewhere in the upper reaches of the Irtysh River Basin in far northeastern Kazakhstan and northwest China, from around Lake Zaysan east to the source of the river in the Mongolian Altai, extending north to the Altai Krai in far southern Russia. This is the area where four countries, Kazakhstan, China, Russia, and Mongolia, all come together.

At some point there would have then been an eastern movement out of the Tobol-Ishim-Irtysh area, possibly following the Irtysh River east towards Lake Zaysan. This movement occurred some time between 425-1000 BCE.

Although the dating is uncertain, Proto-Turkic may have broken up into its initial branches around 400-425 BCE. Those branches were Proto-Kipchak-Karluk, Proto-Yakutic and Proto-Orkhon-Oghuz-Karakhanid. That means that proto-Turkic existed from around 425-1000 BCE, as Bulgaric left around 1000 BCE to head to the west, so Proto-Turkic is dated to 3000 YBP.

Orkhon-Oghuz-Karakhanid

This group is typically placed in the Southern group of Turkic alongside Oghuz, Seljuk and Yugur-Salar, possibly because of a presumed common lineage. However, linguistically, their common

relationship is not very clear, and the relationship of Oghuz-Seljuk to Orkhon-Karakhanid is more indirect than direct.

Certainly there are retentions of archaic Orkhon-Karakhanid forms that can be found in Oghuz-Seljuk (mostly from Karakhanid), but similar forms can also be found in Karluk languages, which obviously also got them from Karakhanid. Further, Proto-Oghuz looks more like a sister to Karakhanid than a descendant like Khalaj. Nor is Proto-Oghuz a direct descendant of Old Uyghur – Yuguric is instead. No doubt, Proto-Oghuz probably split off very early from Old Uyghur rather than being directly descended from it.

Rather than retentions, we should be looking at shared innovations between Orkhon-Karakhanid and Oghuz-Seljuk, and there is little evidence for this. Therefore, the separation of Orkhon-Karakhanid as a separate node from Southern or Oghuz-Seljuk seems to be valid.

One question is whether there is enough resemblance between Orkhon-Karakhanid, Orkhon Old Turkic, Uyghur-Karakhanid, Karakhanid, and Old Uyghur to justify creating a coherent group here, much less an entirely separate node of Macro-Turkic. The available evidence is certainly suggestive, as there are parallels between these languages in lexis and particularly in grammar and phonology.

One objection that has been raised is that there is presently a dearth of material from these ancient languages, and therefore it is impossible to arrive at any conclusive statements about whether they form a coherent group sufficiently different as to be split off from the rest of Macro-Turkic. It is true that we lack a full picture of these languages, however recent publications have devoted entire chapters to a comprehensive look at some of them. Orkhon Old Turkic in particular seems well described.

At any rate, what material we do have seems sufficient to not only collocate these languages into a group but also to split them off as a separate node entirely, if only due to the extreme divergence of their existing descendants, Khalaj and Yuguric.

Orkhon-Oghuz-Kharakanid split off from Proto-Turkic around 300 BCE.

Proto-Oghuz may have then split off from this group very early, perhaps around 250 BCE. Karakhanid then split off from Old Uyghur and Proto-Orkhon Turkic in 100 BCE. Orkhon Old Turkic proper formed around 500 CE.

Orkhon Old Turkic

Commonly called Old Turkic, a better name for this extinct language is **Orkhon Turkic**, the same language as used in the Orkhon carvings. It split off in 500 CE.

Uyghur-Karakhanid

Kharakhanid

Both of the Karakhanid languages, **North Kharakhanid** and **South Karakhanid**, are long since extinct.

Arghu

Khalaj is often said to be a descendant of an ancient language called **Arghu**. Arghu is little known.

Khalaj

Khalaj (Lewis et al 2009i), in reality a macrolanguage, is spoken in Iran by [40,000 people](#) in 47 villages around the religious city of Qom north of Arāk in the Central Province (Knüppel 2010), and was until recently thought to be a dialect of South Azeri. However, in 1968, Gerhard Doerfer showed that Khalaj was a separate language that actually preserved several archaic features not found in any other Turkic languages (Doerfer 1971). Another account was written earlier, the first study of the language (Minorsky 1940). It is now properly seen as a separate branch of Turkic in its own right ([Tekin 1990](#)).

Khalaj is often placed in the Southern Oghuz-Seljuk branch of Turkic, probably due to its origins with the Karakhanid, Orkhon Turkic and Old Uyghur languages. Geographically, this is indeed the southern branch, but there are many reasons to argue against a geographical classification of Turkic. Language family classifications are better done in terms of common features or how archaic the languages are. Khalaj, Yugur and Salar all originate from this grouping, and all are quite different from all other Turkic languages.

Khalaj has been undergoing Persianization for about 20 years, and presently language death is beginning.

It has many odd archaisms that cannot be found in Old Turkic, nor do they exist in Chuvash and Yakut (Doerfer 1989). These forms probably go all the way back to Proto-Turkic.

It is probably best seen as an offshoot of the extinct South Karakhanid language (Mudrak 2002). It may also go back to an ancient language called Arghu (Johanson and Csató 1998). Karakhanid, Old Uyghur and Orkhon Old Turkic were all spoken in the Altai region around this time.

Khalaj has been in Iran a long time, during which a strong Oghuz layer was laid on top of the ancient Karakhanid base. Khalaj is often mistakenly placed with the rest of the Oghuz-Seljuk languages probably due to massive Oghuz borrowing that went in during the long period that they were in Iran. But is this Oghuz layer what is so important about Khalaj? At any rate, since when is language classification done on the basis of borrowings? In that case, English could should be classified as a Romance language due to Latin and French borrowings.

There is an [interesting theory](#) that at one point the Khalaj were in Afghanistan before moving to Iran.

There is good evidence from Pashtun oral literature and sources

written in Dari and Pashto that the Khalaj may be the parent group of the famed Pashtun tribes. The homeland of the Pashtuns is in western Afghanistan in a region called Ghor. Next to Ghor resided the Turkic Khalaj tribes (no doubt speaking Khalaj). The vast majority of the Khalaj apparently lost their identity early, dropped their Turkic language and adopted the Pashtun language (Sierakowska-Dyndo 2013).

The descendants of these Turkic tribes are thought to be the Ghilzai, who, together with the Durrani, make up the two largest Pashtun tribes in Afghanistan (Sierakowska-Dyndo 2013). So the Pashtuns derive in part from Turks, and Turkish traditions, similar to the Pashtuns' own ways of life, were adopted. The other large group, the Durrani, have never accepted the Ghilzai as equals. Instead they see themselves as the only Pashtun tribe in the land. The Durrani were the Pashtuns that existed before the Khalaj become Pashtuns.

All of this must have happened quite some time ago, probably before 1000 CE. A small group of Khalaj who retained their native language at some point, probably long ago, must have moved into Iran. We know that the move from Afghanistan to Iran must have occurred long ago because the huge Oghuz layer in Khalaj indicates a long presence in Iran.

During Minorsky's initial studies 60 years ago, Minorsky, who spoke Turkish, found he was able to understand a fair amount of Khalaj due to its substantial Oghuz overlayer.

Doerfer, who led several expeditions to the Khalaj region around 1970, concluded that Khalaj, which has several dialects, is actually an independent group of [small languages](#) (Knüppel 2009). The dialects show remarkable and wide differences, probably due to the ancient time depth of Khalaj, one of the oldest Turkic languages.

The following are the [dialects](#) of Khalaj (Doerfer 1998 p. 276):

Western Khalaj: Borz-abad and Xalt-abad

Northern Khalaj: Vashqan, Mehr-e Zamin, and Chahak

Northeastern Khalaj: Mushakiya, Espit, Safid-ala, Ahmad-abad, Baghe, Yek, and Salafchegan.

Southern Khalaj: Shana

Central Khalaj: Xarrab

East Central Khalaj: Mouan, Sorxadeh and Mansurabad

Khalaj Proper: Many dialects, 33 different ones.

At the very least, Western Khalaj including the dialects of Borzabad and Xaltabad must be split off as a separate language. Its differences from the rest are so great that even the other Khalaj speakers say that Western Khalaj is not their language (Doerfer 1998 p. 276).

Western Khalaj then is a separate language from Khalaj proper.

Much more data is needed on the Khalaj lects, as there appear to be one or more structurally separate languages in this group (Knüppel 2009).

Old Uyghur

Yuguric (Yugur-Salar)

Yuguric is the name I give to the group that is usually referred to as Yugur-Salar. This group is often placed in the Southern or Orkhon-

Oghuz-Karakhanid family, apparently for geographic reasons, but this classification does not explain the extreme divergence of Yuguric.

Yuguric is typically placed in the Southern branch of Turkic as a separate branch in that group apparently due to its roots in Orkhon-Kharakanid, the same base from which Khalaj springs.

But it should not be placed there (which in its case is not even a geographical classification as Yugur is spoken in northern China) because it is not even partially intelligible with any other Southern languages. Although a lot of this is due to heavy Sinitic, Tibetan and Mongolic layers of borrowings, the truth is that even Yuguric's Turkic words are often barely recognizable due to the extreme sound changes they have undergone. Yuguric has undergone the most extreme sound changes of any Turkic group, mostly due to the borrowings mentioned above.

Yugur lacks any obvious connection to the Oghuz-Seljuk languages, so it is wrong to put it in with them.

Yuguric is as different from the rest of Turkic as Khalaj. Indeed, they share similar origins in Uyghur-Karakhanid - Yuguric originating in Old Uyghur and Khalaj in South Kharakanid. Old Uyghur probably split off from Uyghur-Karakhanid around 100 BCE. In order to illustrate the profound divergence of Yuguric, it should be joined with Khalaj in a separate node away from Turkic proper.

Yugur, Western Yugur, Yellow Yugur, or Saryg Yugur is not the same language as Uyghur. Yugur is a language spoken by 12,297 people in Gansu Province, China. It is spoken from north Menyuen northwest towards Yumen along the Hexi Corridor and the Old Silk Road. Other languages in the area are Mongolian, Chinese and Tibetan. Yugur probably emerged between 700-840 CE in various places along the Silk Road as a sort of a trader's jargon.

Yugur should not be confused with Eastern Yugur or Shira Yugur which

is a Mongolic language. The Shira Yugur are the furthest east group of the Yugur who probably originally spoke a Turkic lect but then switched to Mongolic.

The first theory of Yugur origins is that it is related to Khakas, and indeed it is sometimes said to be most closely related to the Khakas language and is also [grouped](#) with Mrass Shor, Ös and Khakas (Tekin 1990). There are resemblances to Khakas mostly in the grammar and somewhat in the phonology of Yugur, but the lexicon shows little resemblance. Yugur grammar and phonology shows more similarity to Orkhon-Kharakanid than to anything else.

Yugur is often grouped with Khakas because one theory is that Yugur emerged from the Yenisei Kirghiz tribes grouped around Lake Zaysan. This group may have been speaking proto-Yenisei Kyrgyz or even better, proto-Altay-Khakas. Despite this grouping, Yugur probably has almost zero intelligibility with any Khakassian tongue or indeed with the rest of Turkic other than the Salar languages.

The second theory is that Yugur emerged from the Old Orkhon Turkic culture after it was destroyed by the Yenisei Kirghiz tribes in 840 CE. According to this theory, the Yugur might have migrated to Turfan and Ganzhou. Hence, Yugur is most closely related to Old Orkhon Turkic. This theory may also explain the similarities with Yenisei Kirghiz.

The third theory is one promoted by the Yugur people themselves. Yugur people themselves say that they moved from Ganzhou to Turfan after the introduction of Islam (Tenishev 1966). This would mean a movement along the Silk Road from the Kingdom of Kocho where Old Uyghur was spoken to where they are today.

Kocho was a Buddhist-Manicheanist Uyghur civilization that existed during the Idikut Phase of Uyghur culture. It was created around 860 CE where the Old Uyghur language was spoken, and even today, this region is considered to be the center of Uyghur culture. Islam was introduced to this region from 934-960 CE.

This hypothesis also explains the similarities between the name Yugur and Uyghur. Indeed, the furthest eastern state of the Uyghurs was the Ganzhou Kingdom (the state was also known as Yugor State) located in Ganzu Province from 870-1036 CE. The descendants of these people refer to themselves as the Yellow Yugur, which is the very group we are discussing here.

A fourth theory is that Yugur is its own branch of Turkic, one that is poorly understood.

Probably the best theory is the third one, that Yugur is a descendant of the Old Uyghur language. However, there has been some minor Yenisei Kirghiz influence on Yugur via migrants from the north of the Yugur area, which may explain the existence of the first theory and part of the second theory.

Salar

[Salar](#) (Lewis et al 2009n), actually a macrolanguage, is a language spoken by [21,000 people](#) (Dwyer 2001) in Qinghai Province, China, and has its [own separate branch](#) within Turkic (Tekin 2010). This classification will follow Tekin to some extent, but instead of putting Salar in its own branch, we put both Yugur and Salar together as Yuguric into a high-level branch. The Salars have been in China since 1400 CE, and their language has been heavily influenced by Chinese and Tibetan.

Salar is a Turkic classification mystery, as it is typically placed in different groups, and there appears to be no consensus regarding where to put it. The origins of Salar are probably located in the breakup of Karakhanid into Chagatai in the 1300's. Therefore, Salar might best be placed in the Karluk branch with Uyghur and Uzbek.

However, rather than putting Salar in Karluk with Chagatai, the best way to see Salar is that it arose around 1400 CE from a Yugur base along the Silk Road. New merchants speaking Chagatai were doing

business in the area.

Hence, the Yugur (Old Uyghur) base which affected the grammar was overlain with a Chagatai layer which affected the lexicon, the result being a sort of Yugur-Chagatai mixed language which then underwent extensive Sinitic, Mongolic and Tibetic influence, mostly in phonology. Really the best way to view Salar is as a Uyghuric tongue. Yugur is straight up from Old Uyghur, and Chagatai itself came from an Old Uyghur-Karakhanid base. Old Uyghur is the common link between the two inputs to the language.

Salar is so different from other Turkic languages that intelligibility between it and the most of the rest of Turkic (other than Yugur) is probably close to zero. Salar is also close to Uyghur and Yugur, particularly Yugur. It has been dramatically affected by Mandarin and the Dongxiang Mongolic language.

Salar consists of two full languages, Eastern Salar and Western Salar. The Salar languages are surrounded by Mongolic languages such as Santa, Monguor, Baoan and Hezhou in addition to Chinese, a Sinitic language. Tibetan is spoken to the south.

Eastern Salar has been [heavily influenced](#) by Chinese and Tibetan. It is spoken in the center of Gansu Province.

Western Salar has been influenced more by Uyghur and Kazakh than by Sino-Tibetan. As a result, the two Salar varieties are [not intelligible](#) with each other (Dwyer 2001); hence Western Salar and Eastern Salar are separate languages. Western Salar is spoken in the far northwest of Qinghai Province from the towns of Linxia in Gansu west towards the city of Xining in Qinghai.

Western Salar has some asymmetrical intelligibility with Uyghur due to significant Uyghur influence. Interviews with Western Salar speakers in Urumqi, Ghulja, Yili, and Qinghai and with Uyghur speakers who have traveled to the Western Salar speaking area have

uncovered lopsided intelligibility between Western Salar and Uyghur (Dwyer 2015).

Western Salar is completely unintelligible to Uyghur speakers, but Uyghur is partly intelligible to Western Salar speakers (Dwyer 2015). This lopsided intelligibility strongly suggests significant bilingual learning of Uyghur on the part of Western Salar speakers who are likely to have learned some of Uyghur, a major language in the region and suggests that Uyghur speakers have not bothered to learn Western Salar, a minor language in the region.

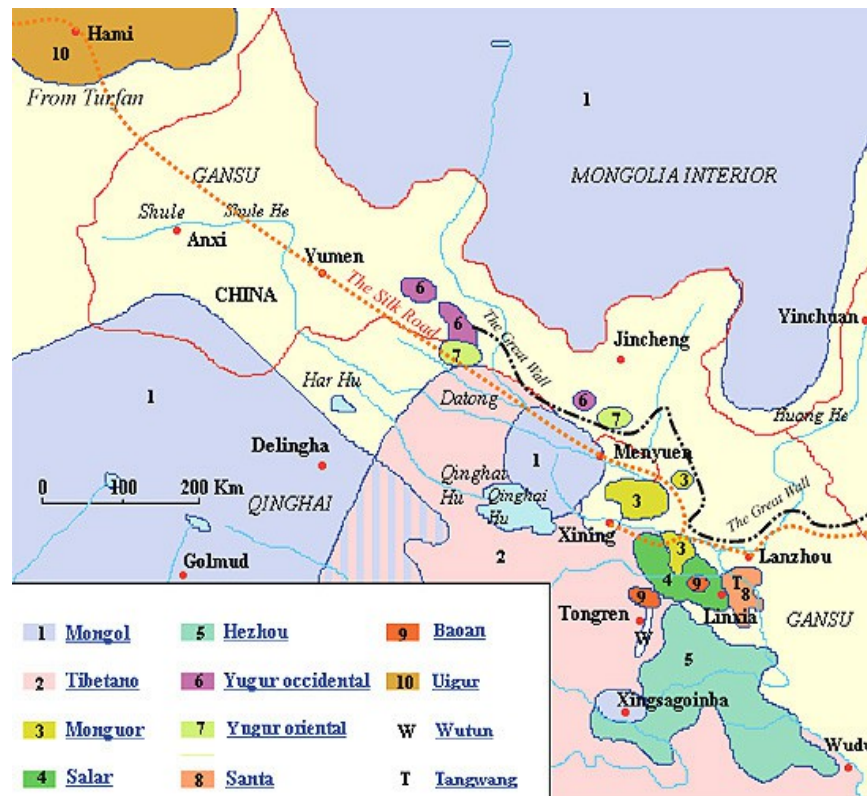


Figure 2: The location of the highly divergent Western or Sarigh Yugur(6) and Eastern Salar languages (4) in China is shown here. Eastern or Shera Yugur (7) is actually a Mongolic language. The other languages are Mongolic and Tibetan.

Yakutic (Sakhaic)

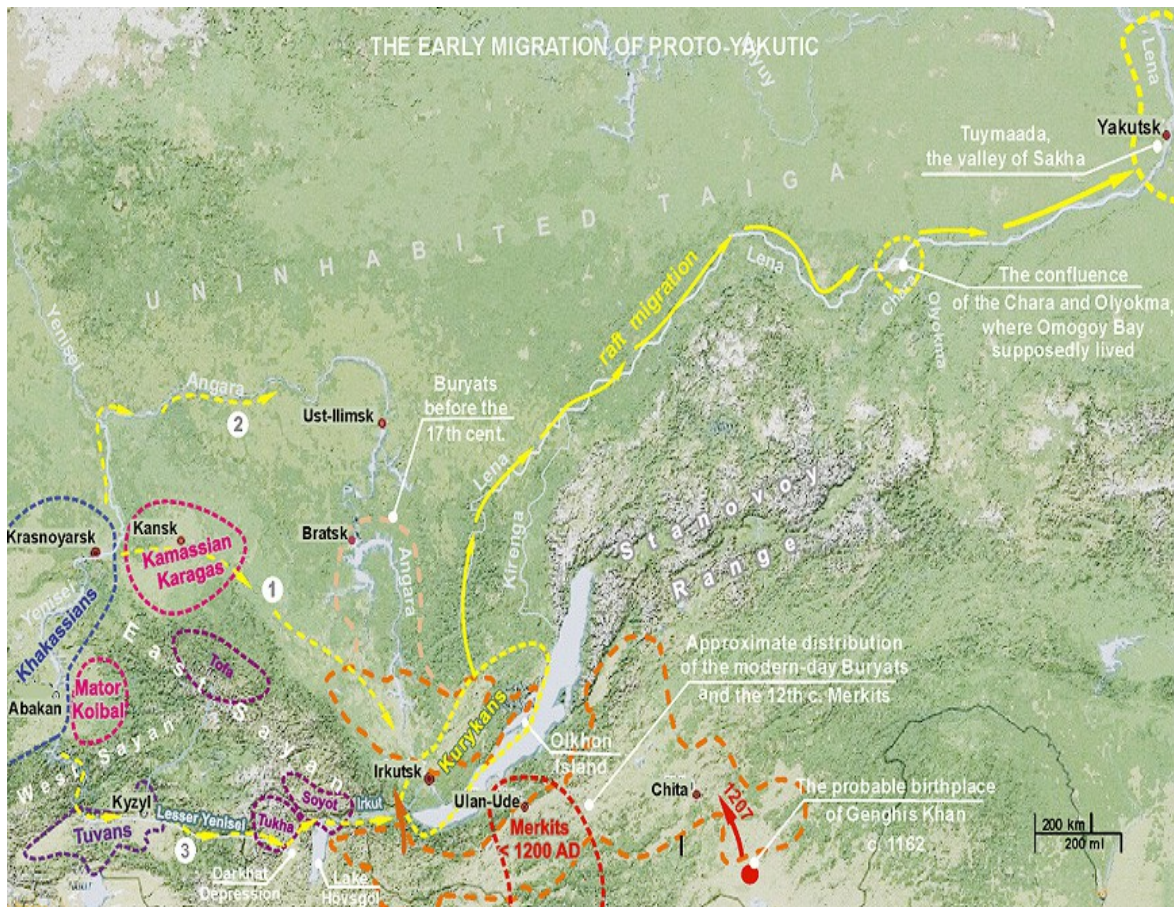


Figure 3: Map showing three possible routes taken by proto-Yakuts from the Urheimat in the Sayans, first to Lake Baikal as the Kuryakans and later to the Lena River. The route labeled (3) may be the most accurate one. Also shows some other languages and their locations at various historical times and the present.

Yakutic is best put into a separate node alongside with Khalaj/Yuguric. Yakutic does not have any clear or obvious relationship with any other Turkic tongue. It has typically been thrown in with a grouping called Siberian which is more of a geographic grouping than anything else. The suggestion then is that Yakutic shares some Siberianisms in phonology, grammar and lexis with the

rest of the grouping.

However, a close look at the so-called Siberianisms reveals that nearly all of them are more properly seen as retentions or archaisms rather than true Siberian innovations. Some of the phonological Siberianisms look more like Mongolic or areal Sprachbund (Mongolic, Sinitic and Tibetic) influences. There is perhaps a single Siberian innovation that would remain after all of the old words have been accounted for.

In recent years, a case has been made for connecting Yakutic and Tuvan or Tuvaic (Stachowski 2011). Although there are surely similarities between the two, especially between Yakutic and Tofa, these appear to be more indirect than direct – meaning that Tofa may have developed under a Yakutic substratum.

However, there are some minor resemblances between Yenisei Kirghiz and Yakutic. This looks more indirect relationship than a direct one. What might have happened was that Proto-Yakutic was an early influence on early Siberian languages or possibly that some or more Yenisei Kirghiz languages developed on a Proto-Yakutic substratum.

Various ages have been given for Proto-Yakutic, one of the first languages to split off Proto-Turkic. These dates range from 200-900 BCE. The best analysis shows that the later dates are better than the earlier ones. A good date for Proto-Yakutic is 300 BCE. The location is hard to pinpoint, but a good guess might be high in the West Sayan Range along the Yenisei River near the present town Ust-Ula very close to Khakassia. In recent years, South Samoyedic languages like Mator, Koibal and Kamass were spoken in this area.

At some point, the Proto-Yakut began moving south down the Yenisei past Kyzyl to the Lesser Yenisei on Yenisei Steppe near the Mongolian border, slowly following the river through the Darkhut Depression towards Lake Hovsagol through the modern territory of the Soyot and Dukha.

When this journey was undertaken is not known, but it was probably complete by 100 CE. They remained in this area for 400 years. Soyot and Dukha may have been affected by Yakutic substrate. In fact, the name Dukha looks remarkably like Yakut. By 500 CE, they were undergoing another movement east towards the western shore of Lake Baikal, where they were known from Chinese sources as the Kuryakans from 500-900 CE.

Yakut has [453,288 speakers](#) and is one of the official languages of the Republic of Sakha in Siberia (Moseley 2010k). Yakut, Tuvaic, and Khakassian languages have 0% intelligibility with the rest of Turkic (Eker 2013). In particular, Yakutic has 0% intelligibility not only with the rest of Turkic but also with the rest of the Siberian Turkic language group in which it is erroneously placed. The reason for this is Yakutic's strong Yeniseian, Mongolic, and Russian influences.

For instance, the closest language to Yakutic is Tuva, but the intelligibility between the two is 0%. 26% of Yakut vocabulary is made up of borrowings from Mongolic and 33% of Russian borrowings (Ratloff 1908).

The specific Mongolian languages are thought to be Middle Mongolian from 1200-1300 continuing to 1400-1600 and Middle Buryat probably from the same period (Kaluzynski 1962). Proto-Mongolian as reconstructed does not look a lot different from Middle Mongolian (Pakendorf 2007). The heavy Yeniseian influence is probably because Yakut developed from a Yeniseian substrate.

There is some Tungusic influence in Yakut, mostly from Evenki, but it is controversial. Some sources say there is little Tungusic influence on Yakut, and there resemblances that exist are typically Yakut borrowings into Evenki and not the other way around.

This theory underestimates the Tungusic influence on Yakut, which is considerable, most of it from Evenki but also some from Even (Anderson 1998). The main influences on Yakut are Mongolic and then

later Russian. Tungusic and Yeniseian are seen as lesser influences, possibly based on Tungusic and Yeniseian substrates in Yakut.

The Mongolic borrowings are somewhat mysterious. Mongolic speakers began moving into the Lake Baikal region in the 1000's in response to the expanding Khitan Empire in Mongolia. Yakut speakers, known at that time as Kuryakans, resided on the southwest side of the lake from 500-1000 CE. The Mongolic group that moved to Baikal, while they lived next to the Yakut for around 200 years, does not look to be the source of much of the heavy Mongolic borrowing in Yakut.

However, starting in the early 1200's and possibly continuing into the 1300's, the Yakut began moving north out of the Baikal area towards the Lena River due to Mongol campaigns against the nearby Turkic Yenisei Kirghiz and others. The campaign against the Yenisei Kirghiz occurred from 1206-1209 (Pakendorf 2007). Not all of them left the area, and the Yakut that remained as Kurykans transformed into the Buryats, a Turkic group that took up a Mongolic tongue.

Although the main movement occurred over 200 years later, some small Yakut groups had already started moving up the Lena by the year 1000 CE (Pakendorf 2007). The main Mongolian expansion outside of Mongolia began only in 1206-1259 CE during the Mongol Empire. This coincides with the beginnings of the Mongolic borrowings into Yakut. This still leaves the major part of Mongolic and Buryat borrowings into Yakut from 1200-1600 poorly explained.

Dolgan (Lewis et al 2009d), [spoken](#) by 4,865 people (Tishkov 2009) in Dudinka and Khatanga counties in former Taymyr (Dolgan and Nenets) Autonomous District and in Anabar County in the Republic of Sakha (Yakutia), (Moseley 2010a) is sometimes said to be a [dialect of Yakut](#) (Tekin 2010).

This was the standard view until the 1950's, when Turkic scholars decided that Dolgan was a separate language. Experts say that this is a different language from Yakut, and it is [not mutually intelligible](#) with

Yakut (UCLA Language Materials Project a). Furthermore, Dolgan speakers refuse to use Yakut learning materials because they are too different from Dolgan (Eker 2013).

The last word on the matter appears to come from Pakendorf, who did fieldwork for her PhD among the Yakut. According to her, there is low MI between Dolgan and Yakut due to Dolgan innovations, phonetic changes, changed semantics for shared lexis and significant borrowing from Evenki. It qualifies as a separate language on both sociopolitical (the Dolgans see themselves as speaking a separate language) and structural grounds (Pakendorf 2007).

However, there are claims that Dolgan speakers have excellent intelligibility with Yakut speakers in the northwestern part of the Yakut zone (Northwestern Yakut). The reason for this may be the heavy Evenki influence on Northwestern Yakut (Pakendorf 2007).

Dolgan differs considerably from Yakut due to [strong influences](#) from the Evenki (Tungusic) language (Humphreys and Mits 1991c). Dolgan was originally a Yakut dialect called Vilyuy, but it split off recently (around 1650-1700 CE) and transformed itself into a separate language (Stakowski 2011) and [underwent independent development](#) (Johnson 2011).

The best theory is that Dolgan are Evenki speakers from the Dolgan and other tribes who switched to Yakut (Anderson 1998). Yakut and Dolgan now form their [own branch](#) of Turkic (Tekin 1990). On this matter, Tekin is certainly correct. I have followed Tekin in this treatment by splitting off Yakutic as one of two nodes below Macro-Turkic and above Common Turkic.

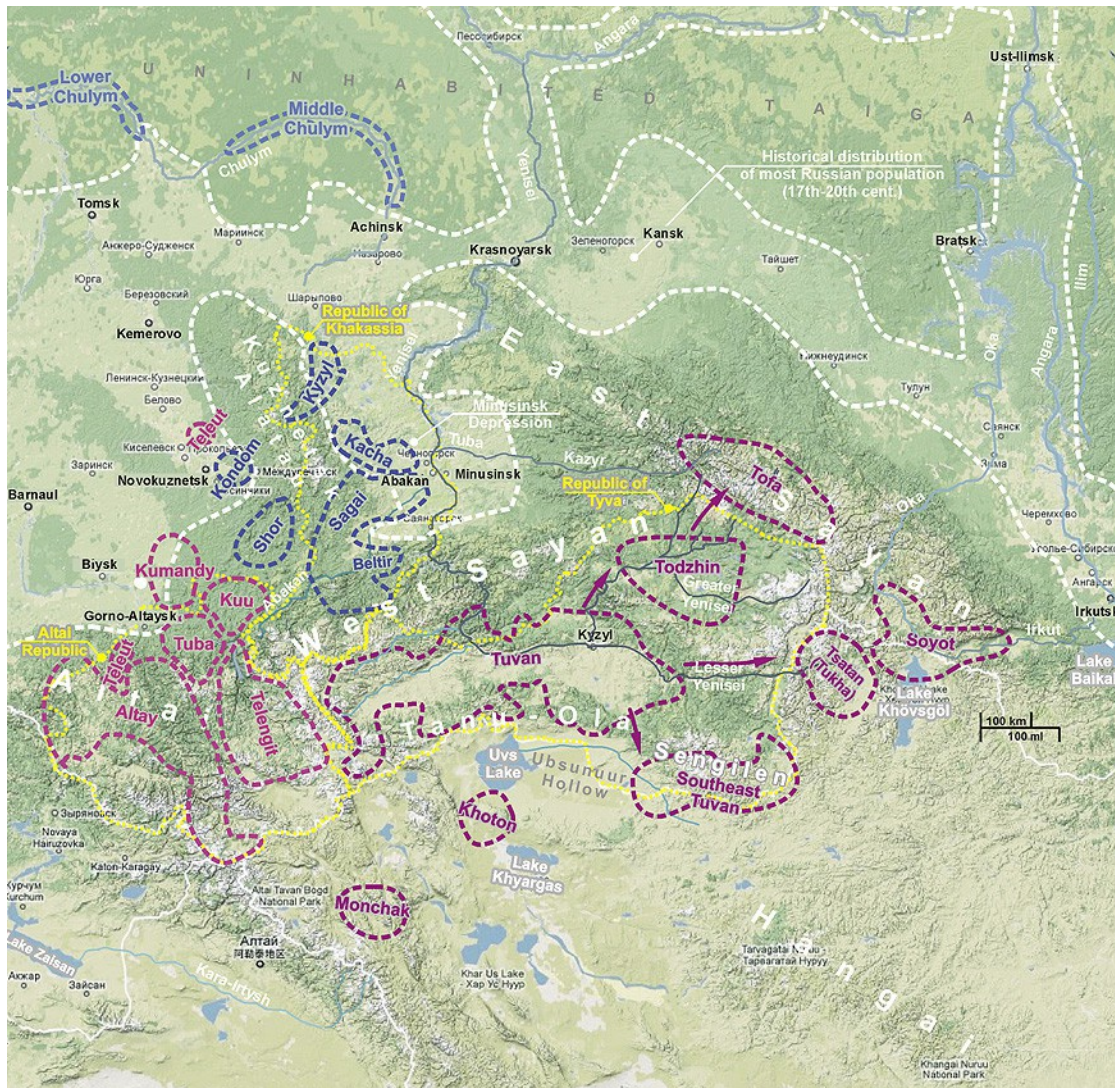


Figure 4: Map showing the Siberian Turkic languages.

Turkic (Common Turkic)

Central (Yenisei Kirghiz or Siberian)

Tuvan (Tuvaic)

Tofa and Tuvan, spoken by 252,474 people (Tishkov 2009) in Siberia, form [a separate branch](#) of Turkic close to Altay in Tekin's classification (Tekin 2010). They are not intelligible with each other, but there are said to be intelligible dialects linking them. Intelligibility between the two may be as high as 85% going by [glottochronology](#) (Dybo 2006), but reported intelligibility is much lower (Eriksson 2013).

Each one has a separate written form. Tofa and Tuvan differ in [other ways](#); for instance, each one has a nominal case that the other one lacks, and conditionals also differ between the two languages, with Tofa retaining an older form (Anderson 2004). Tuvan has borrowed heavily from Mongolic and recently from Russian.

The entire group of Tuvan languages is referred to as the Sayan languages. They are split into two groups: *Steppe Sayan*, encompassing Tuvan and Tuha, and *Taiga Sayan*, encompassing Dukha, Tofa, Soyot, and Todzhin.

The Tuvans were ruled by Mongolics prior to 200 CE. It is not known what language they spoken at this time. Around 500 CE, they were conquered by Turkics. Proto-Tuvan probably goes back to around 250 CE. The location may be in the Lesser Yenisei region around the Yenisei Steppes where Soyot and Dukha are now spoken all the way over to Lake Khövsgöl.

Steppe Sayan

Steppe Sayan consists of two languages: Tuvan and Tuha.

At least part of Steppe Sayan was influenced by Yeniseian, as these people used to speak the Kol' language but were assimilated to Tuvan. However, we do not know which part of Tuvan was affected by Yeniseian.

The Western Tuvan lects have been [heavily influenced](#) by the Southern Altay Telengit dialect (Dougherty undated).

[Tuha, Tuhalar, Toha, Uighur Urianghai or Uriankhay](#), the furthest east of the Sayan lects, is spoken by a small group of only 20-30 members of the Soyot tribe, all elderly, with some passive speakers in their 50's.

It is spoken in and around the villages of Tsagaan-Üür and Üürin-Tsagaan in the Tunka Valley in the far north of Mongolia in the Khövsgöl Aymag district to the east of Lake Khövsgöl from the shores of the lake north to the border with Buryatia (Eriksonas 2012). They are [cattle-herders](#) in contrast to the Dukha on the other side of the lake who are reindeer herders (Ragagnin 2012).

One theory is that, like the Tofa, they were originally speakers of the Samoyedic language Mator who [switched](#) over to a Turkic tongue (Eriksonas 2012).

[Another theory](#), one that I follow in this treatment, is that is Tuha is a Steppe Sayan language, albeit with some features setting it aside from the rest of Steppe Sayan. This theory also states that instead of being Turkicized Samoyedics, the Tuha at base are the descendants of ancient Kurykan Turks (early Yakut) from the 700's who [did not move north and get assimilated](#) to the Buryats like most of their contemporaries (Ragagnin 2012). At the moment, the latter theory seems to be preferred.

The Tuha moved into the region around [1600-1650 CE](#) (Eriksonas 2012). This is apparently a group of Soyot who split off from the rest of Soyot when the others moved north into Buryatia long ago. Since Soyot and Tuha have been separated for ~400 years, there is probably difficult intelligibility between them.

Tuha pronouns were mostly borrowed from Mongolic and differ dramatically even from Dukha pronouns. The Tuha have a harder time

speaking to the Tuva than other Steppe Sayan do.

[For instance](#), the Tuva cannot understand Tuva textbooks, while the Monchak can (Eriksonas 2012). 30% of Tuva words are from Mongolic.

Tuva may not be intelligible with any other Tuvan language.

Tuva has a variety of sometimes divergent dialects that are sometimes considered to be separate languages. However, on close examination, all of these are more properly seen as Tuvan dialects rather than full languages.

Altai Tuvan, Altai Sayan, Tuwa, or Menggu is a Tuvan dialect spoken by people in [three different countries](#), in the Bayan Ölgei (Olgii) and Khovd Aimak region of western Mongolia (the Mongolian Altai) as Gök Monjaq, Aq Soyan and Xara Soyan and across the border by in Xinjiang Province in three villages near Kanas Lake (Johanson 2001). The speaker [population](#) totals less than ~3,500 people -~1,850 in Mongolia and 1,700 people in China (Atwood 2004).

Altai Tuvan has heavy Mongolian influence. It has been split off from Tuvan for since 1730 (Mongush 2012b), and the dialect is quite different from Tuva (Johanson 2001). Recent investigators [characterize](#) Altai Tuvan as a separate language (Living Tongues 2007a). It is also referred to as a highly divergent dialect of Tuvan.

However, reports indicate that Altai Tuvan, at least the Jungar Tuvan variety, is [completely intelligible](#) with Tuvan (Todoriki 2010), hence Jungar Tuvan is a dialect of Tuvan and not a separate language.

Dzungarian Tuvan or Jungar Tuvan, [also known as](#) Khövsgöl Tuvan, Kubsugul Uriangkhai, Khövsgöl Tuvan, or Tuxa (Janhunen and Salminen 1993), is spoken by [1,700](#) speakers (Atwood 2004) in the northernmost part of Sinkiang in the Uyghur Republic of China near

the borders of Kazakhstan and Mongolia in the Chinese Altai. They live in the [villages](#) of Akkaba in Habahe/Kaba County, and Kanas, and Kom-Kanas in Burqin County (Atwood 2004). The Jungar Tuvans split off from the Altai Tuvans in Mongolia and moved into China between 1700-1800 CE. Their language has many Kazakh and Mongolian words layered onto a Tuvan base.

Jungar Tuvans in China understand radio broadcasts in Tuvan proper well. Reports say that intelligibility is [still excellent](#) between Jungar Tuvan and Tuvan Proper (Todoriki 2010). In addition, Jungar Tuvan texts are easy to read from a Tuvan perspective (Anderson 2013).

Jungar Tuvan in China and Altai Tuvan in Mongolia were formerly part of one dialect, but they became separate around 1900 and have since undergone separate development, the form in Mongolia undergoing heavy Mongolian influence and the form in China coming under heavy Chinese influence (Dougherty undated). In recent years, communication has become more difficult between Jungar Tuvans and Mongolian Tuvans (until recently part of one group) due to massive Chinese borrowings going into Jungar Tuvan (Mongush 2012b).

This dialect is simply Altai Tuvan as spoken in China.

Tsengel Tuvan is the name of the one of the Altai Tuvan dialects spoken in Mongolia. It has recently been characterized by David Harrison as a [separate language](#) (Harrison undated b) spoken by 1,500 people in far northwestern Mongolia near the borders of Russia and China. Tsengel is a district of Bayan-Ölgii Province in western Mongolia. The capital of Tsengel is Khushoot. This dialect is spoken in the village of Tsengel Sum in Bayan-Ölgii province.

Tsengel Tuvan is simply one of the two Altai Tuvan dialects spoken in Mongolia along with Monchak.

Tsengel Tuvans either [cannot](#) or will not use Tuvan language materials (Bavuu-Syuryun and Salchak 2012); nevertheless, [classes](#) are

somehow held in Tuvan in this area (Atwood 2004). However, the Tsengels' rejection of Tuvan materials seems to have more to do with social and ethnic factors than linguistic ones (Anderson 2013).

The divergence between Tsengel Tuvan and the rest of Steppe Tuvan may be great enough to split it off as a separate language (Anderson 2013). However, Tsengel may have good understanding of Tuvan since that is what David Harrison spoke with them while he was studying them (Anderson 2013). Tsengel Tuvan appears to be a dialect of Tuvan, but this remains somewhat up in the air.

Monchak, Kök-Monchak, Monjak, or Mondzhak is the name of another Altai Tuvan dialect in Mongolia. K. David Harrison states that Monchak and Tsengel Tuvan are two separate languages, and that [Monchak](#) (Harrison undated a) is a moribund language spoken by [less than 150 people](#) (Harrison 2007) in the [village](#) of Buyant in Khovd province, Western Mongolia (Atwood 2004).

The Monchaks lived in the Eevi River region of Xinjiang Province, China until they [moved](#) into Mongolia around 1930 (Harrison 2007), so this is simply a late movement of Jungar Tuvan speakers from China into Mongolia. Harrison's rationale for stating that Monchak and Tsengel Tuvan are separate languages is not known. In contrast to the Tsengel Tuvans, the Monchak report that they are able to use Tuva textbooks (Eriksonas 2012). However, there is [no Tuvan language instruction](#) in the Monchak region (Atwood 2004).

Expert opinion at the moment holds that Monchak and Jungar Tuvan at least are not separate languages in their own right but instead are ethnolects or sociolects, or in other words, dialects, of Tuvan Proper (Anderson 2013).

Hovd or Kovd is a second Mongolian Tuvan dialect reported by Harrison (Harrison 2006), apparently also subsumed under Altai Tuvan. Kovd is a town 80 miles southwest of the Tsengel region near Khar-Us Lake. Based on what can be gleaned from the meager data

on Internet sites, Hovd appears to be [the same language](#) as Monchak (Enduring Voices 2010).

On the other hand, there are also 6,000 speakers of apparently a more or less normal variety of the Tuvan language living in the Hovd region (Johnson 2001), so Hovd may refer to this group.

Khoton, characterized by some as a separate language, is said to be yet another Altai Tuvan spoken in Mongolia.

Khoton apparently refers to a group called the [Tsaatan Khoton](#) (Wind of Mongolia 2011). This group has its origins in a Soyot or more probably Tuha group that was taken prisoner by Mongolian speakers. They decided while in captivity to abandon Turkic in favor of a Mongolian language.

Khoton simply means “Turk” in Mongolic. They presently [speak Buryat](#) (Tatár 1979).

In this treatment, I have subsumed the Monchak dialect in Mongolia and the Jungar Tuvan and Tsengel Tuvan dialects in China under the umbrella of a single dialect, Altai Tuvan. Altai Tuvan is a Steppe Sayan dialect consisting of three subdialects, Tsengel Tuvan, Monchak and Jungar.

Taiga Sayan

The ancestors of Taiga Sayan and Todzhin speakers arrived in the region as Proto-Samoyedic speakers ~2,000 BCE. Around the same time, Evenki (Tungusic) and Yeniseian speakers moved into the Tuvan area. Soon afterward, Turkic speakers (probably proto-Yakuts) moved up from the south and began to influence these Samoyedic, Yeniseian and Tungusic groups.

Taiga Sayan is composed of at least four different languages – Soyot,

Tofa, Dukha, and Todzhin.

Southeast Tuvan or Kyzyl Tuvan may be a separate language according some scholars. However, it could be intelligible with Todzhin. Until this is cleared up, SE Tuvan will remain an unclassified Taiga Sayan lect.

Kyzyl Tuvan is a [Taiga Sayan variety](#) spoken around Lake Tere-Khöl southeast of the city of Xyzyl in the far southeast of Tuva near the Mongolian border (Ragagnin 2009). Preliminary analysis suggests that the Tere-Khöl dialects are similar to the northeastern group containing Todzhin and may be a dialect of the Todzhin language (Mongush and Sat. 1967). Tofa speakers cannot understand Kyzyl Tuvan (Eriksson 2013).

Tofa, Tofalar or Karagas, [spoken](#) (Salminen 2007a) south of the city of Nizhneudinsk and in the villages of Alygdzher, Nerkha and Verkhnyaya Gutara in Nizhneudinsk County in Irkutsk Province, has a number of archaic features that are not present in Tuvan and also has heavy [Mongolian influences](#) (Rassadin 2005b). The name Karagas should be discouraged as it more properly refers to the extinct South Samoyedic group by that name.

Tofa split from Todzhin/Tuva in ~1750.

Tofa was formed by layering Tuva onto a [Samoyedic base](#) (Janhunen and Salminen 1993). The fact that there are quite a few similarities between Tofa and Yakut (Rassadin 1978) implies that Yakut may also have been part of the Tofa substratum, or perhaps they were Yakut speakers who switched to a Tuvan lect.

Tofa is presently moribund, with only 35 speakers, all over age 40 (Anderson and Harrison 2003). Tofa is currently undergoing obsolescence as a language, and the vowel harmony system in particular is starting to [fall apart](#) (Anderson 2004).

Recent investigators [characterize](#) Tofa as a separate language. However, one source says that Tofa, Dukha and Todzhin are largely mutually intelligible (Harrison 2002). Nevertheless, recent reports say that Tofa has difficult intelligibility [not only](#) with Tuva but also with the closely related Todzhin (Mongush 2012a).

According to Valentin Rassadin, Tofa is not mutually intelligible with Todzhin even though there is only a single mountain ridge separating them. Rassadin also said that a Tofa man once went to the Xyzyl region of Tuva where they speak the Kyzyl Tuvan dialect (close to Todzhin), and the Tofa speaker could not understand Kyzyl Tuvan at all (Eriksonas 2013).

Of course the Tofa have probably never even ventured down to Lake Hövsgöl to even meet the Dukha speakers, much less speak to them (Eriksonas 2013). Tofa had [difficult intelligibility](#) of Soyot (Donahoe 2003). Recent intelligibility testing has shown that Tofa is indeed a separate language (Anderson 2013).

Todzhin, Todzhu or Northeastern Tuvan, is spoken by 4,406 people (Tishkov 2009) in the Tuva region. Todzhin is more properly placed in the [Taiga Sayan](#) group (Rassadin 2013) than in the Steppe Sayan group. There has long been a suggestion that the Tofa and the Todzhin are the same people, but this is not the case.

The Todzhin appear to be a group of Tuva who have [assimilated](#) to the Tofa group and have taken up reindeer herding (Donahoe 2003).

They were recognized as separate peoples (probably speaking separate languages) even prior to [1700 CE](#) (Donahoe 2003). Traditionally placed in Steppe Sayan, Todzhin is better placed in Taiga Sayan. The reason Todzhin is so different is due to the fact that they used to speak a Samoyedic tongue – the Mator, Koibal, Kamas, and Karagas were all in the region - and were assimilated to Turkic.

Recent investigators feel that Todzhin is a [separate language](#) (Living

Tongues 2007a). Tuva speakers refer to Todzhin as a [separate language](#) (Donahoe 2003). A report said that Tuvans [could hardly understand](#) the “harsh and loud” Todzhin language at all (Kosterin and Zaika 2004).

Nevertheless, Todzhin, Tofa and Dukha are said to be largely mutually intelligible (Harrison 2002). However, a Tuva magazine reported that even the closely related Tofa and Todzhin had [difficult intelligibility](#) (Mongush 2012a). Valentin Rossadin states that Tofa and Todzhin are not mutually intelligible (Eriksonas 2013).

In recent years, Todzhin has been used less and less, and the Todzhins increasingly use Literary Tuvan.

Soyot was spoken in the [Oka and Tuka regions](#) of Buryatia in Russia (Rassadin 2005b). The center of their territory was southwest of the city of southwest of the city of Irkutsk just north of the Mongolian border. The Soyots, like the rest of Taiga Sayan, apparently originally spoke a Samoyedic language, but they were Turkicized at some point. One source says the Turkicization may have occurred [between 600-700 CE](#) or possibly even earlier (Rassadin 2013).

Between 1600-1650 CE, the Soyots left the Lake Khövsgöl region of Mongolia and moved north to the Buryatia area. The ones that stayed behind probably became the Tuha. Another account says that around 1650 CE, Soyot became separated from (Greater) Dukha. The Samoyedic group went to Soyot (and probably went north out of Mongolia), and the Turkic group went to Dukha-Tuha (and probably stayed in Mongolia).

Soyot was still widely spoken up until the 1920’s, but then the group began abandoning it for Buryat. Soyot in Buryatia is apparently [extinct](#) (Todoriki 2010). It probably went extinct sometime after the early 1990’s when it was [still spoken](#) by a few old men (Rassadin 2005b). The group now speaks a dialect of Buryat that has been [creolized](#) from a Soyot base (Todoriki 2010). Nevertheless, there are

still some elderly people who remember quite a few words.

As Soyot split from Tofa long ago and hence pursued an independent track of development, the Tofa had a hard time understanding Soyot, which they referred to as a separate language. Todzhin intelligibility of Soyot is unknown, but if the Tofa had a hard time of it, the Todzhin no doubt did too, as these two languages are very close.

Tsaatan or Dukha, [also known as](#) Mongolian Reindeer Tuvan or Caatan (Salminen 2007a), is spoken by about 250 speakers west of Lake Khövsgöl in the mountainous western half of Khövsgöl Aimak in the far northernmost part of Mongolia near the border with the Tuva region and Buryatia. This is a part of a group that split off from the Todzhin as they were shifting from Samoyedic to Turkic. An already Turkicized group went to the Dukha and a group that was still Samoyedic but later shifted to Turkic went to the Soyots.

The Greater Dukha made up [two different groups](#) (Donahoe 2003). One group, the northeastern group, has been in the area since 1600-1650 CE, and another group, the southwestern group, only left the Tuva region in the 1940's. It [appears](#) that the northeastern group is actually the Tuva, and the southwestern group is the Dukha.

Some of the Dukha also came from the Tere-Khöl region in the 1940's (Ragagnin 2006).

Dukha has a [very close](#) relationship with Todzhin, differing mostly only in a number of Mongolian loans (Donahoe 2003). Since their [arrival](#) in the region, Dukha has undergone heavy Mongolianization (Ragagnin 2006). Dukha also has several features that differ from Tuvan.

Dukha is [not intelligible](#) with Tuva or any other Tuvan language (Fitzhugh 2002) and is therefore a separate language. However, other researchers report that Dukha is largely mutually intelligible with Tofa and Todzhin (Harrison 2002). Yet Dukha is part of a dialect chain with Tofa, which implies that they are separate languages. Dukha is also

very close to Soyot.

Mongolian researchers created an alphabet for Dukha that has also [proved useful](#) for Tofa since they have similar sound systems (Rassadin 2013). This shows the close relationships of these two languages.

Dukha is definitely a separate language within Greater Tuvan (Anderson 2013).

Khakassian

Khakas (Lewis et al 2009h), a macrolanguage, is in [its own branch](#) of Turkic in Tekin's classification (Tekin 2010). In this classification, Mrass Shor is together with Ös, Khakas and Yugur in a single group ([Tekin 1990](#)). Shor is sometimes said to be a dialect of Khakas.

Yet intelligibility between the two may only be ~55% going by [glottochronology](#) (Dybo 2006).

Khakas is [spoken](#) by 42,600 people (Anderson 2004) [in](#) the western half of the Minusinsk Steppe region on the upper Yenisei in southern Siberia, mainly in the Republic of Khakassia but also in Uzhur and Sharypovo counties in Krasnoyarsk Region and in adjacent parts of the Republic of Tuva in the Russian Federation (Moseley 2010i). Turkish has [zero intelligibility](#) of Khakas (Tekin 1978).

Recent work indicates that Khakas is made up of a complex dialect continuum (Anderson and Harrison 2006, Anderson 2013). Dialects of Khakas include Abakan, Beltir, Kachin (Kaca, Khaas, Xaas), Koibal (Xoibal), Xyzyl (Khyzyzl, Xyzyl), Sagai (Sagaj, Saghai), Yarin and Shor.

Xyzyl is unintelligible with the rest of Khakas, [as far apart](#) from the rest of Khakas as Ös is (Living Tongues 2007a). Therefore, Xyzyl is a

[separate language](#) (Anderson 2012). Xyzyl is spoken along the upper reaches of the Ijus (Üüs) River in northern Khakassia and Krasnoyarsk Kray. Xyzyl has a [very close](#) relationship with the Melet dialect of Ös (Anderson 2004); indeed, Xyzyl Khakas is actually transitional to Melet or Upper Ös (Anderson 2013).

Intelligibility tests showed that Xyzyl failed intelligibility tests with the Shor and Sagai Khakas dialects, but did better when tested with Xaas. Xyzyl speakers reject the Khakas label for their language. The verbal system in Xyzyl is very different from all of the rest of Khakas. (Anderson 2013).

All in all, there seems to be a good basis for splitting Xyzyl off from the rest of Khakas, though the evidence is somewhat equivocal (Anderson 2013).

Kamas or Kamas Turk is the source of much confusion. Kamass, often listed as a dialect of Khakas, is actually a Samoyedic language that only went extinct in the past few decades. A group of Kamass speakers abandoned their Samoyedic language for Khakas some time ago, and therein lies the confused belief that Kamas is a Khakas dialect.

The truth is that one language, [Samoyedic Kamas](#), died in 1989 when Klavdia Plotnikova died. It was spoken in the eastern part of the Minusinsk region in what is now the Krasnoyarsk Region. The language became gradually restricted to the Kan and Mana River Basins and their sources on the far northwestern slopes of the Eastern Sayan Mountains. The last Kamas-speaking community lived in the village of Abalakovo (Moseley 2010e).

The Turkic Kamas lect of Khakas then went [extinct](#) in 2010 (Moseley 2010f).

There is some confusion between the Shor language proper and the Shor dialect of Khakas.

Shor Khakas is spoken by descendants of Mmass Shor speakers who moved across the Kuznetskiy Alatau to the Minusinsk region into the Khakas region between 1500-1800 CE. It has a lot of Shor elements, and this is a Shor dialect somewhat restructured into a Khakas dialect.

Shor Khakas is apparently a dialect of Khakas and not a separate language (Anderson 2013).

Koibal has a similar origin with Kamas Khakas.

It represents another group of Samoyedic speakers who abandoned their Samoyedic lect for Turkic in recent centuries (Anderson 2013). Koibal had a large Samoyedic substrate. It was spoken in Abalakovo village in the Krasnoyarsk Region alongside Samoyedic Kamass.

There are now reports that it has [gone extinct](#) (Moseley 2010f). Koibal was sparsely attested so it is hard to make judgments about it, but it seems that it was a dialect of Khakas and not a separate language (Anderson 2013).

Koibal and Kamas were the two dialects of the Kamas Samoyedic language. It seems clear that when many Kamas speakers shifted to Turkic, speakers of the two dialects carried their names over to Khakas, as Koibal Kamass became Koibal Khakas and Kamas Kamass became Kamas Khakas, with both of them probably subsumed under the label Kamas Turkic.

Yarin were [former speakers](#) of the Yeniseian language Yarin or possibly the more commonly known Arin which went [extinct](#) in the 1730's (Anderson 2005) as speakers [shifted](#) to Khakas (Georg 2003).

Sagai, along with Xaas, is the basis for *Literary Khakas*, which is an artificial language not much different from actual Sagai. Sagai is presently expanding at the expense of all of the other lects (Anderson

2005). Sagai is best seen as a Khakas dialect and not a separate language (Anderson 2013).

Xaas, Kacha or Kachin is sometimes seen as a separate language, but a better analysis is that it is a Khakas dialect (Anderson 2013).

Beltir is sometimes seen as a separate language.

This group formerly spoke Tuvan but shifted to Khakas around 1800 (Anderson 2005). There is a Tuvan substrate in this dialect, but it also is best characterized as a dialect of Khakas (Anderson 2013).

Speakers of all Khakas lects have traditionally referred to all of the other Khakas varieties as “languages” (Anderson 2013). That does not seem to be enough to split them all off into separate languages, but it does seem to imply significant divergence in the lects.

Khakas then consists of one language, Literary Khakas, under which a complex dialect continuum is subsumed, and another language, Xyzyl, which is a structurally separate language.

Fuyü Gïrgïs, also known as Manchurian Kirghiz, is not related to Kirghiz as its name suggests; instead it is more [closely related](#) to Khakas, Chulym and Yugur (Johnson 2011). The name Kirghiz probably derives from the ethnonym of the speakers of the Siberian languages as Yenisei Kirghiz people.

This language is spoken in Heliojang County in northeastern China in and around Fuyu County, Qihar, 180 miles north of Harbin (Hu and Imart 1987).

It was originally an outlying dialect of Khakas whose speakers split off and moved to China from Russia possibly around 1700 CE, but it is now best seen as a [separate language](#). In 1980, it was spoken by a

majority of adults in about 100 homes in the town, but since then, most adults have switched over to Mongolian and most of the children are now speaking Mandarin Chinese that is taught at school (Hu and Imart 1987).

It is spoken by only 10 elderly people out of a group of 730 people (Salminen 2007a). In truth, Fuyü Gïrgïs is moribund, there are no more full speakers, and the 10 speakers are better characterized as semi-speakers or rememberers who recall some words (Dwyer 2015).

Actual intelligibility between Fuyü Gïrgïs and Khakas is unknown, however, Khakas may have had significant intelligibility (except for some Mongolian loans) of Fuyü Gïrgïs when it was a full language (Dwyer 2015). There is probably no good way to determine Khakas-Fuyü Gïrgïs intelligibility now that Fuyü Gïrgïs is moribund. For the moment, it seems best to leave Fuyü Gïrgïs as a full Turkic language rather than a Khakas dialect.

Shor

Shor, actually a macrolanguage, is sometimes thought to be a [dialect of Khakas](#) (Tekin 2010), but this is may be a confusion between the *Shor dialect of Khakas* and Shor proper, which according to *Ethnologue* is a separate language.

Indeed, [glottochronology](#) indicates that intelligibility between Shor and Khakas may only be ~55% (Dybo 2006). Nevertheless, Shor, Khakas, and Xyzyl speakers can communicate face to face in a somewhat difficult and muddled way (Anderson 2013), but the same can be said of Spanish and Portuguese speakers (intelligibility 54%) and Spanish and Italian speakers (intelligibility ~35%), so that does not mean much in the end.

Shor is [spoken](#) by less than 1,000 people (Salminen 2007a) in Siberia [along the river valleys on the southwestern slopes](#) of the Kuznetskiy Alatau Range in Kemerovo Province (Anderson 2004).

The Shors were formed between 1500-1600 CE by a mixing of Ugric, Samoyedic, Yeniseian and Turkic groups in Mountain Shoria south of the Kemerovo region (Stukova 2006). The Shors briefly had their own district called Shoria, but it was eliminated in 1939, and the region later developed into one of the largest industrial areas in the USSR, the Kuzbass Region.

Actually, there are [two different dialects](#) of Shor, and their differences are [great](#) (Janhunen and Salminen 1993, p. 44); so great, in fact, that they can be seen as separate languages.

Southern or Kondoma Shor (spoken along the Kondoma River) may be close to Northern Altay and is actually a separate language.

Northern or Mrass Shor (spoken along the Mrass River) is close to Khakas and is also a separate language.

There are [claims](#) that Mrass and Kondoma are more or less mutually intelligible (Ager 2008-2013). At any rate, presently dialect leveling is occurring in all of Shor.

However, the conclusion of a symposium in conjunction with the publication of *Shorica: A Shor Electronic Database* in July 2001 was that Kondoma and Mrass are very different lects that typologically show features of two separate languages which have only recently begun to merge into a single literary language (Stukova 2006).

But this merging process has been slowed by the deep differences between the two languages. Therefore, Shor (Literary Shor) is a pluricentric language that constitutes two separate languages – Kondoma and Mrass – within it.

One classification has Mrass [grouped](#) with Ös, Khakas and Yugur, while Kondoma is grouped with Lower Chulym and the Kumandin and Chelkan Northern Altay languages (Tekin 1990). The intelligibility of

Kondoma with Northern Altay and of Mrass with Khakas is not known, but Mrass and Sagai Khakas appear to be close.

However, [glottochronology](#) indicates that Shor itself may have only 60% intelligibility with Khakas and ~40% with the Altay group (Dybo 2006), so it does not look like either Shor language is part of either the Altay or Khakas language group.

The Shor macrolanguage is presently [moribund](#) (Anderson 2004).

More [recent scholarship](#) has cast doubt on the connection of Kondoma with Northern Altay and instead connects it with the Southern Altay Teleut language. On the eastern edge of the Kondoma region, Kondoma gives way to Mrass, which then transitions to the Sagai dialect of Khakas (Funk 2005a).

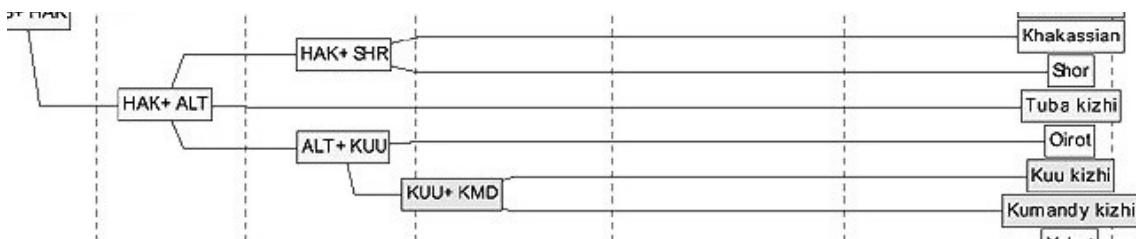


Figure 3: Glottochronology for Khakas, Shor, and the Altay languages. This chart shows that Khakas and Shor are quite separate and that Kumandin is a separate language from the rest of Northern Altay. From [Dybo 2006](#).

Ös (Chulym)

Chulym or *Ös* is actually a macrolanguage, often said to be a dialect of Khakas, whereas Lower Chulym is said to be a [dialect](#) of Northern Altay (Tekin 1990). Chulym is indeed very close to Khakas although

they are separate languages – certain Upper Chulym dialects are transitional to certain Khakas lects (Anderson 2013).

Middle and Lower Chulym are [spoken](#) on the middle part of the Chulym River, a tributary to the upper Ob', mainly in Tegul'det County in Tomsk Province and Tyukhtet County in the Krasnoyarsk Region north of the Altai Mountains in the Ob River Basin near the city of Novokuznetsk (Salminen 2007a).

Chulym or Ös is made up of [two dialects](#), Middle Chulym or Ös and Upper Chulym, which are mutually intelligible (Anderson and Harrison 2006, Anderson 2013). The difference between Upper and Middle Chulym is less than that between British and American English (Anderson 2013). Upper Chulym is closer to Khakas.

Nevertheless, the *Melet or Middle Chulym* and the *Tutal or Upper Chulym* dialects, spoken in two different villages – Melet and Tutal (Anderson 2013) - show evidence of having been separated for quite some time, 200-300 years.

Upper or Tutal Ös is [transitional](#) to Xyzyl Khakas (Anderson 2013).

Ös [split off](#) (Janhunen and Salminen 1993) from Khakas between 1700-1800 CE, and the language has since undergone independent development with Ös being influenced by Selkup (a Samoyedic language) and Siberian Tatar with additional influences from Ob-Ugric and Yeniseian.

Actually the best analysis is that Ös speakers were mostly Yeniseian (probably Pumpokol) speakers and partly speakers of the Uralic Ob-Ugric Khanty language (Anderson 2013) until 1750 CE when they abandoned their Yeniseian tongue for a Turkic language.

Ös is often [grouped](#) with Mrass Shor, Khakas and Yugur (Tekin 1990). *Ethnologue* says that Ös is very close to Shor, and the two may even

be the same language. Others feel that Shor and Ös are the same language as Khakas. However, Ös speakers cannot understand Khakas, so they cannot be part of one language.

However, [K. David Harrison](#), who has studied the language extensively, says that attempts to lump Ös in with both Khakas and Shor are incorrect, and it is distinct from either of them (Anderson and Harrison 2006).

Ös then is a separate language. Ös has only 20 speakers remaining, all over 60 years old.

Lower Chulym deserves its own *Ethnologue* entry according to Harrison, which implies that he does not believe it is a dialect of Northern Altay. Harrison states that the closest language to Ös is actually Lower Chulym. This means that Lower Chulym is not as close to Northern Altay as had been theorized.

In addition, Russian missionaries to the Altai in the 1830's said that Lower Chulym speakers could not understand Northern Altay and vice versa (Kharlampovich et al 2001). Lower Chulym is different enough from Ös such that there is [difficult intelligibility](#) between the two (Beaumont 2013). In addition, Ös speakers say that Lower Chulym written materials are [unrecognizable](#) (Living Tongues 2007b), meaning they are two separate languages.

Lower Chulym is commonly grouped with Kumandin, Chelkan (Northern Altai) and Kondoma Shor.

Lower Chulym was heavily [Tatarized or Tatar-like](#) when last recorded (Living Tongues 2007b, Anderson 2013). In 2006, Chulym was reported to be extinct (Anderson and Harrison 2006). Another report stated that it went extinct in 2011.

However, recent indications are that Lower Chulym is not yet extinct,

and there are still a few speakers left (Anderson 2013).

Küärik is said to be a third major dialect of Chulym, spoken along the lower course of the Kiya River, a tributary of the Chulym. Some observers say it was a separate language, but [the truth is](#) that this was a dialect of Lower Chulym (Anderson and Harrison 2004, Anderson 2013). At any rate, it went extinct around 1900, but not before it was [documented](#) by V. V. Radlov around 1850 (Funk 2005b).

There were a number of tribes who spoke what were apparently Chulym lects that have since gone extinct. These include the *Yachin*, *Kumysh*, *Shuy* and *Kamlar* lects formerly spoken in the former Kuznetsk district of Tomskaya Province. All of these groups were [Russified](#) at some time in the 1800's, quit speaking their languages and started speaking Russian instead (Funk 2005b).

Altay

Altay, a macrolanguage, is spoken in the Altai region of Russia where China, Russia and Mongolia all come together. Altay is split into [Northern Altay](#) (Lewis et al 2009l) and [Southern Altay](#) (Lewis et al 2009o), and intelligibility between the two is poor.

Intelligibility between the two may be 55% going by [glottochronology](#) (Dybo 2006). In Tekin's newer classification of Turkic, Northern and Southern Altay form two [completely separate related](#) branches (Tekin 2010). All of Altay has [less than 25,000 speakers](#) (Anderson 2004). Altay has been [influenced](#) by Mongolian and Samoyedic.

Turkish intelligibility of Altay is [extremely low](#), close to zero (Tekin 1978).

However, Northern Altay is a macrolanguage, not even a separate language; instead, it is made up of three languages: Kumandin, Chelkan and Tuba. Northern Altay is spoken by 4,000 people in the northeastern river valleys of the Altay region in Siberia.

The Northern Altay languages Kumandin and Chelkan, together with Lower Chulym and Kondoma Shor, now form a [grouping](#) (Tekin 1990). All of Northern Altay derives from Turkicized Nenets and Yeniseian speakers, and [references](#) to the group go back as far as 600-800 CE (Potapov 1969).

Southern Altay, a macrolanguage, is spoken by ~20,000 people in the central and southwestern river valleys of the Altai region of Siberia, and is composed of two languages: Oirot Altay and Teleut, with Telengit as a Teleut dialect.

Reports indicate that there is sometimes [difficult intelligibility](#) among Cheklans, Kumandins, Teleuts, Telengits, and Tubalars (Galiullina 2011). This implies that all of these are separate languages.

However, we lack data on the intelligibility of Altay Kizhi and Teleut/Telengit. Teleut at least is probably [not fully intelligible](#) with Altay Kizhi (Potapov 1969). There is a suggestion that the Telengits may have a hard time communicating with the Altay Kizhi, but it is not proven yet.

Altay is better seen as a geographical grouping rather than a linguistic one; some taxa included under Altay may be better placed elsewhere, and some lects now outside of Altay may actually be better put into Altay (Anderson 2013).

Indeed, no linguist who has looked at the data has ever felt that Altay represented a single taxon, and multiple splits in Altay beyond Northern and Southern are slowly gaining acceptance (Anderson 2013).

An alternative view is that both Southern Altay and Northern Altay consist of dialects that are clearly distinct from each other, but are nevertheless used in communication with speakers of other dialects (but, obviously, not with speakers of the other, fully distinct languages also referred to as Altay). In other words, speakers of Southern Altay

languages communicate with each other via their own languages as do speakers of Northern Altay languages with other Northern Altay speakers, however, communication between Northern and Southern Altay is not possible. (Salminen 2014).

However, in the future we shall probably see more splits in Altay, as Anderson suggests.

For one thing, linguists have already stated that Kumandin and Teleut are structurally separate languages, Kumandin being separate from Northern Altay and Teleut being separate from Southern Altay (Bitkeeva 2005, Funk 2005a). Since Telengit may be a dialect of Teleut (Anderson 2013), if Teleut is separate, then Altai Kizhi, being the only Southern Altai lect left, must be a separate language also.

If Kumandin is separate, then that leaves a Northern Altay containing Tubalar and Chelkan. However, Tubalar probably split from the rest of Northern Altai in 700-800 CE, 1200-1300 YBP (Potapov 1969). It is closer to Khakas and Shor than to the rest of Northern Altay, and glottochronology indicates that intelligibility between Tubalar and the rest of Northern Altay must be poor (Dybo 2006).

That leaves us with only Chelkan, which is then a separate language by default. So as we can see, based on the judgments of current linguists working with these lects, there are already at least five languages in Altay, which is exactly what was determined in the current treatment.

Northern Altay

Kumandin (Bitkeeva 2005) is spoken by 1,044 people (Tishkov 2009) in Krasnogorskoye, Solton, Kytmanovo and Tselinnoye counties in the Altai Region and in adjacent counties in the Republic of Altay as well as in the towns of Tashtagol and Sheregesh in Kemerovo Province. The 1,044 figure is [too low](#) (Moseley 2010j); the actual number is probably ~5,000.

It is generally thought to be a [Northern Altay dialect](#) (Tekin 2010), but it is actually an independent language closely related to Shor and Khakas. The reason for advocating a split is apparently on [structural grounds](#) (Bitkeeva 2005). On phonetic grounds, Kumandin is close to both Khakas and Shor.

Friar Makarii Glukarev, a Russian missionary to the Altai from 1829-1844 that the Kumandins could not understand Southern Altay text and vice versa (Kharlampovich et al 2001). Kumandin is [grouped with](#) Chelkan, Lower Chulym and Kondoma Shor (Tekin 1990). Based on [lexicostatistics](#), Kumandin may have 80% intelligibility with Altay Kizhi or Chelkan (Dybo 2006).

[Chelkan or Kuu-Kizhi](#) (Tazranova 2005) is another [Northern Altay dialect](#) (Tekin 2010) that is actually a separate language. It is [grouped with](#) Lower Chulym, Kumandin and Kondoma Shor (Tekin 1990). Chelkan is [spoken](#) by 539 people (Tishkov 2009) in the villages of Kurmach-Baygol, Suranash, Malyy Chibechen' and Itkuch in Turochak County (Salminen 2007a).

The figure of 539 speakers is presumably [too low](#) (Moseley 2010j); the actual figure is probably ~2,500. Based on [glottochronology](#), Chelkan may have ~80% intelligibility of Kumandin and as such is a separate language (Dybo 2006). The Chelkans probably split off and become a separate group around [1625 CE](#) (Potapov 1969). At one time, they had a close relationship with the Kondoma Shors.

[Tuba, Tubalar, or Tuu-Kizhi](#) (Humphreys and Mits 1991a) is another Northern Altay dialect that is in truth a separate language. Tuba is spoken by 439 people (Tishkov 2009) [in the region of](#) Turochak, Choya, and Mayma counties in the Republic of Altay in the Russian Federation (Salminen 2007a).

The 439 figure is presumably [too low](#) (Moseley 2010j); the actual figure is probably ~2,500. It is very far from the rest of Northern Altay.

In fact, it is closer to Khakas and Shor than to the rest of Northern Altay and also further from the rest of Northern Altay than Southern Altay itself based on glottochronology (Dybo 2006).

Looking at [glottochronology](#), intelligibility of Tuba with the rest of Northern Altay is probably low, on the order of 40% or less (Dybo 2006). References to the Tuba may go back as far as [700-800 CE](#) (Potapov 1969), which would make sense in light of the fact that it is the most divergent of all Altay languages.

Southern Altay

The Southern Altay group is split into three different groups – the Altay Kizhi, the Teleuts, and the Telengits.

This group may have split up around 1750 CE after their parent group, the Teleses, was [attacked](#) by a Kazakh feudal lord named Kochkorbai in the 1750's. They sought protection and received it from the Russians. Another view is that they go back further, as the Altay Kizhi and Teleuts are listed as separate groups [as far back](#) as the 1200's (Potapov 1969).

References to the Telengit as a separate group go all the way back to the 700's. It appears that Southern Altai may go back 1,000 years or more.

[**Southern Altay Proper, Altay Kizhi or Oirod Altay**](#) (Janhunen and Salminen 1993) is spoken by about 6,000 people in the central and southwestern valleys of the Altai Region. This is apparently a separate language from Teleut.

The intelligibility of Altay Kizhi with Telengit is [not known](#) (Galiullina 2011).

This language is the basis for the Standard Altay language, which has been taught in schools and used in the media for 70 years now. The best term for this language may be Altay Kizhi. The term “Oiroṭ” is confusing and best avoided, as all of Altay was called Oiroṭ until the 1940's.

Teleut (Janhunen and Salminen 1993) classed as a [Southern Altay dialect](#) (Tekin 2010), is a separate language altogether according to *Ethnologue*.

Russian missionaries to the Altai in the 1830's said that Teleut and other Southern Altay speakers could not understand each other's texts (Kharlampovich et al 2001). Teleut speakers [cannot use](#) materials written in the Altay Kizhi language (Nevskaya 2005).

Recently, scholars studying Teleut have [decided](#) that instead of being an Altay dialect, it is more properly seen as an independent language with a particularly close relationship with the Kondoma Shor language (Funk 2005a).

Teleut is [spoken](#) (Salminen 2007a) by 1,892 people (Tishkov 2009) mostly in the western Kemerovo region. The homeland of the Teleuts is in Shebalino County in the Republic of Altay, but many were dispersed between 1600-1800 CE. Now [the bulk of Teleut speakers](#) are located north of the actual Altay region in the Bachat region in Belovo, Gur'yevsk and Novokuznetsk counties in the western part of Kemerovo Province and in the Chumysh region of Tselinnoye County in the east of Altay Province (Salminen 2007a).

Telengit is a Southern Altay dialect spoken in the Altai region by 2,314 people (Tishkov 2009). Intelligibility with Altay Kizhi is [unknown](#) (Galiullina 2011) but is probably difficult. The Telengits may be relative newcomers to the region [having come recently](#) from the Tuvan country (Potapov 1969). Indeed, the Western dialect of Tuvan is similar to Telengit. They may have originally been Tuva speakers who switched to an Altai dialect.

The Telengits may have only split off from the Teleuts around [1600 CE](#) (Potapov 1969). However, [other reports](#) say that the Telengits split off from the Steppe Altay around 700 CE and moved to the high mountains where they have stayed raising sheep ever since (Mote 2004). Intelligibility data for Telengit and Altay Kizhi needed.

However, there is a possibility that Telengit is simply a dialect of the Teleut language.

A local Teleut who is also a good linguist believes that Telengit is merely a dialect of Teleut. Intelligibility between Telengit and Teleut has not yet been tested, even anecdotally (Anderson 2013). Arguing against the notion that Telengit is a Teleut dialect is the suggestion that Teleut and Telengit may experience [difficult intelligibility](#) (Galiullina 2011) with each other.

For the last 200 years, a [small diaspora group](#) of Southern Altai, supposedly Teleuts, composing only 50 people, has been living in the Altai region of Sinkiang Province, China (Salminen 2007a). However, there are claims that the Southern Altai dialect that they speak is similar to Telengit, not Teleut. They are said to speak a Telengit-like dialect that has come under the influence of other languages in China. They may still retain their language.

Karluk-Kipchak

The Karluk-Kipchak grouping generally inhabits the Great Steppe Region of Eurasia stretching all the way from the area west of the Tian Shan near Lake Balkash in the Zhetsu Region of southeastern Kazakhstan all the way to Ukraine or even Lithuania.

The Karluk-Kipchak Homeland is probably around the Upper Irtysh River (Black Irtysh) area in the Altai Mountains. This area encompasses the region from the source of the river in the Mongolian Altai Range to Lake Zaysan in northeastern Kazakhstan. Karluk-Kipchak split off from Proto-Turkic around 300 BCE, and the Karluk

languages themselves may have split off from Karluk-Kipchak very soon afterwards.

Karluk

The Proto-Karluk homeland existed near the Altai Range before 600 CE. There was a later movement down to the Tian Shan Range between 650-800 CE. Karluk began to break up soon after the southward movement, perhaps as early as 730 CE.

Chagatai

Chagatai was an ancient Turkic language that was a very popular and important Turkic koine in Central Asia during the Medieval Period. It was formed by an old Uyghur-Karakhanid base combined with later Karluk and Kipchak influences. This language formed after the Mongols invaded the region they inhabited in the 1200's. The era of Classical Chagatai was in the 1500's.

It was particularly widely spoken in Afghanistan. Whether Chagatai could best be described as Uyghur or Uzbek is uncertain. Some sources refer to Classical Chagatai as Uyghur, and indeed it was spoken by members of the Uyghur ethnic group. Others refer to Chagatai as Old Uzbek. Under the Soviet classification, Chagatai was treated as Old Uzbek, and Chagatai was used as part of the base for the creation of literary Uzbek around the time that Chagatai went extinct.

It is usually thought to have gone extinct in 1921.

However, in 1965, a man named László Szimonisz sent a letter to Gerhard Doerfer reporting that Chagatai was [still spoken](#) in an area south of Tehran where he had stayed recently (Doerfer 1969, p. 13). The name or exact location of the group was not given, and the claim remains unproven.

Turcologists say that even if a small group of speakers exist, this is not particularly important as the language for all intents and purposes died as a true literary language in 1921. One expert said that finding a small group of Chagatai speakers would be akin to finding some speakers of Ottoman Turkish at this late date in Turkey (Suer 2014). As far as we are concerned here, Chagatai is extinct.

Nevertheless, Chagatai is used as one of the languages of instruction for children of the Dāyī ethnic group living around the Andkhōy Oasis and in the neighboring districts of Qaramqōl and Arabgul in the far northwest Afghanistan (Baldauf 2007b). There may a long tradition of Uzbeks in Afghanistan using Chagatai as a written language (Baldauf 2007a).

Considering that the lexicon of Dāyī Turkic is heavily Southern Uzbek, all members of the Dāyī ethnic group speak Southern Uzbek (Baldauf 2007b) and Chagatai could be seen as Old Uzbek, Dāyī children should be able to understand instruction in Chagatai well. In addition, Chagatai is still widely studied by students in Turkey.

Uyghur or New Uyghur (a macrolanguage) and Uzbek (a macrolanguage) are fairly close, but they are still probably only ~75% intelligible. Uyghur is spoken in Sinkiang Province, China.

There is huge dialect diversity in Uyghur, some believe it represents [more than one language](#) (Hahn 1998), and in fact, Uyghur does appear to be more than one language

There is excellent intelligibility (over 90%) between all Uyghur dialects (Imin 2013, Dwyer 2015). Lop Nur is the exception, as it is actually a separate language.

MI between Turkish and Uyghur is low, ~20% (Tekin 1978).

[Ili Turki](#) (Lewis et al 2009e) is a dialect of Uyghur, traditionally characterized as a separate language close to Uzbek, [spoken](#) in the eastern parts of the Ili Valley in the Uyghur Autonomous Region in

Sinkiang, China. It is said to be distinct from Uyghur itself (Salminen 2007a). However, according to Adrienne Dwyer, Ili Turki is indistinguishable from the Central Uyghur spoken in the Ili Gujia region of Xinjiang (Lewis et al 2009e).

Ili Turki speakers arrived in Sinkiang from the Fergana Valley in Uzbekistan and Kyrgyzstan around 1800. At the time it was thought to be linguistically close to Uzbek, and traditionally, the dialect was often characterized as an outlying dialect of Uzbek with heavy influence from other Turkic [languages](#). However, Ili Turki has apparently assimilated heavily to Uyghur in the 200 years since speakers moved from the Uzbek area. Hence, Ili Turki is best seen as an outlying Uzbek dialect that moved from the Fergana Valley to Xinjiang 200 years ago and has since undergone such Uyghur influence via Central Ili Gujia Uyghur that it is now a Uyghur dialect.

It is [spoken](#) by only 30 or so families, and the younger speakers are shifting to Kazakh or Standard Uyghur.

According to *Ethnologue*, Ili Turki is a [mixture](#) between Chagatai (Proto-Uzbek-Uyghur) and a Kipchak dialect of Uzbek (Grimes 2000), possibly Kwarezm-Kipchak Oghuz, that has since undergone heavy Uyghur influence to the point where it is now a dialect of Central Uyghur.

Lop Nur was traditionally thought to be a Uyghur dialect, is possibly better considered to be a separate language. The present characterization of Lop Nur as a Uyghur dialect has more to do with politics than science as both Uyghurs and the Chinese state have a strong political motivation to call Lop Nur a Uyghur dialect. Lop Nur is not particularly close to Standard Uyghur; in fact, it is as close (or closer to) Kirghiz as is to Uyghur (Dwyer 2015).

The Lop Nur or Lopliks are an ethnic group who live near the marshes in the northeast part of the semi-dry lake in the Tarim Basin (Stein 1912).

They have traditionally live along the lower stretches of the Tarim River where it emptied into the the Tarim Basin into several lakes called Karakoshun (which the Chinese refer to by its Mongolian name Lop Nur) that shifted as the Tarim shifted its course and ended in different places (Stein 1912).

Actually, Karakoshun is the name of the lakes in the Tarim Delta while Lop Nur is the name for another of the lakes.

They were able to fish along the riverine poplar and Toghrak forests of the Lower Tarim and in the marshy lakes of the River Delta. They sporadically sowed fields with both barley and oats when the weather permitted and lived in mud brick homes. They also killed wild camels for their meat (Stein 1912).

Already by 1906 the old way of life along the riparian forests of the Lower Tarim and the lakes of the Tarim Basin was beginning to be lost as the Lop Nur moved northward to new settlements as far as Karashar. At this time, the Lop Nur had already abandoned their old nomadic way of live in the marshes for a more settled life of sedentary farming along the Lower Tarim for 1-2 generations (Stein 1912).

In 1906, the Lop Nurs were living chiefly at the settlements of Charklik and Abdal. In Charklik, they had settled into farming life, and at Abdal, the last remains of Lop Nur traditional hunting and fishing life with reed huts remained, to which wheat farming had recently been added. The traditional way of life utilized camels for transportation and wool for clothing, carpets and other things. Clothing was sheepskins, cotton "Chapans," felt socks and leather "Charuks." Winters were extremely cold; the area resembled a frozen desert at that time of year and the rivers were covered with ice (Stein 1912).

At this time, heavy Tamarisk growth was encroaching on the fields to the north of Charklik due to desiccation Fields in this area were much

drier, being cultivated only every 2-3 years. Further north, heavy sand dunes were covered with dead Tamarisk in a sandy jungle as the area grew drier. North of Abdal was a salt-covered steppe with briny lagoons full of reed beds and large shallow lakes that filled and dried up depending on the year. Further north were salt pools, dry salt-covered lake beds and then an old riverine jungle, now dead and extremely eroded and swept with harsh winds, icy and in the Arctic-like winter (Stein 1912).

The ancient capital of the Lop Nor, possibly named Yamen, was located even further to the north, in an area with eroded sand dunes, icy winds and an occasional salty oasis. It was probably a location on the old Silk Road (Stein 1912).

The lake was very large in 1928, reaching 1,200 square miles. In 1952 and 1972, human intervention, including a large reservoir, made conditions in the Lop Nur region much worse as the Tarim River tributaries that fed the lakes in the area began to be diverted upstream for agriculture, and large dams were constructed, causing further and serious desiccation. This resulted in the drying out of all of the lakes and the death of the extensive poplar and Tamarisk forests along the lower course of the river. Extensive cutting of the forests for firewood only worsened matters. Beginning in 2000 in an effort to reclaim the ecosystem, water was diverted back into the system and a forest restoration project was begun.

The Lop Nur region was originally referred to as Loulan or the Loulan Kingdom and later Shanshan, a Chinese tributary state from 55 CE, from ancient times of the Earlier and Later Han Dynasties until the T'ang Dynasty. The residents were speakers of the Indo-European language Tocharian. Charklik was the main settlement of the Loulan region. The Loulan region was located just to the west of the Chinese border at that time. During the Han Dynasties, this was the area from which Chinese Western expansion first originated (Stein 1912).

From 300's-400's CE, it was controlled by speakers of the ancient Indian language Gandhari Prakrit, written in a script called Kharoṣṭhī,

who were led by a Maharaja.

This era was related to immigration from Gandara in the Kushan Empire in what was then called northwestern India (now the Northwest Territories of Pakistan and far eastern Afghanistan) who moved into the area in 200 CE. They also brought Buddhism to the region. From the 400's-600's, the area was gradually abandoned due to attacks by nearby nomadic states. From the 700's-800's CE, the area was controlled by Tibetans. The Tibetans lost the area in the middle of the 800's CE.

They are different from the Uyghur ethnically and anthropologically, that is, they look different from the Uyghur. Instead of looking like the Iranid Caucasian-Mongoloid mixed Uyghur, they look Mongolian (Stein 1912). Their physical features resemble the Kirghiz, which is interesting as Lop Nur is close to the Kirghiz language. One account says that the Lop Nur are simply the original Uyghurs who never mixed in with the Iranian Caucasoids who came after 1300. Therefore the original Uyghurs were Mongolic types, which is in accord with the historical data.

Lop Nur is nearly dead. Its condition is similar to Fuyü Gïrgïs. Lop Nur is moribund with only a few semi-speakers remaining. It was probably last spoken as a full language in the 1960's. Excellent Lop Nur recordings were made at this time, but they were thought to have been lost, destroyed in the Cultural Revolution. However, they have just recently surfaced and should be a valuable addition to Uyghur studies.

From a century ago, we have excellent evidence that Lop Nur was not the same language as Uyghur. In 1906, Lop Nur was not inherently intelligible to Uyghur speakers from Yarkant and Khotan (Stein 1912). It is dubious that Lop Nur became intelligible to Uyghur speakers in the sixty years between 1908 and the 1960's. The fact that Lop Nur was not intelligible to Uyghur 100 years ago provides cautious evidence for the characterization of Lop Nur as a separate language.

At any rate, besides mutual intelligibility, a good case could probably be made for splitting Lop Nur on structural grounds. If Lop Nur really is closer to Kirghiz than it is to Uyghur, it makes no sense to call it a Uyghur dialect.

Uzbek is actually a macrolanguage. Uzbek is spoken in Uzbekistan and Afghanistan, with some outlying speakers in China. MI between Turkish and Uzbek is similar to Turkish and Uyghur, ~10% (Tekin 1978). Uzbek and Kazakh are not fully intelligible with each other (MI may be ~75%), but there is an intelligible dialect between them.

Äynu, Aini, Aynu, Ainu, Eyni or Abdal is a mysterious language spoken in the Xinjiang region by a nomadic ethnic group called the Äynu. Äynu is spoken only by the males of the ethnic group. The females all speak Uyghur. The language is a sort of a secret tongue. Äynu males speak Äynu when around Uyghurs so the Uyghurs cannot understand them, but they speak Uyghur at home and to Äynu females who do not speak Äynu.

The Äynu people are very much disliked by the Uyghur for unknown reasons, and there is much discrimination against them.

Äynu is an odd language. Many people say it is not even Turkic at all, others say it is Turkic with a heavy layer of Iranian vocabulary, and others say that it is Iranian with Uyghur grammar. It has been commonly characterized as a mixed language, and if so, it is not even Turkic at all, as mixed tongues are outside of known language families since they are so mixed between two different major language families that it is uncertain which major group they are primarily a part of.

Some are so thoroughly mixed together that classification attempts fail and the language is simply characterized as a mixed language outside of any known language family. The theory of Äynu as a mixed language was investigated but found to be wanting.

Probably the best characterization of Äynu is of a Uyghuric language with Iranian vocabulary (Johanson 2001).

Further, there is controversy about whether Turkic layer is Uyghur or Yugur.

Uzbek is a macrolanguage spoken by millions of people mostly in Uzbekistan. Turkish intelligibility of Uzbek may be low, around 20% if we base it on Tekin's analysis of Turkish-Uyghur intelligibility, as Uzbek and Uyghur are very similar.

Northern Uzbek, spoken in Uzbekistan, is a separate language (Lewis et al 2009m). It has many Russian loans.

Northern Uzbek is spoken in Chimkent and to the north of Tashkent. The Tashkent dialect is the basis for the standardized language (Sengupta 2003).

Southern Uzbek, spoken in Afghanistan, is a separate language (Lewis et al 2009q). Southern Uzbek has many Persian loans. There is fair intelligibility between two two Uzbek languages, but there are significant differences in the grammar.

Kipchak

The Kipchak languages are all descended from the Kimak Confederation which ruled the area around Lake Zaysan. They came under heavy influence of Proto-Oghuz-Karakhanid-Uyghur languages in this area from 100-400 BCE. The Kipchak languages were connected with the era of the Golden Horde between 1200-1500 CE. During this period, these dialects were collectively known as Middle Kipchak (Johanson and Csato 1998).

Most of them probably did not begin to split off until later around the 1500's or even later which explains the high degree of multiple intelligibility among them. The classic formulation of Kipchak omits

Nogay and includes Kazakh. Both are incorrect. Nogay is very much a Kipchak language, and Kazakh is a Karluk language that is often wrongly placed in this group.

Cuman

Certain Kipchak lects are best seen as the descendants of the ancient Cuman Kipchak tongue spoken in the Middle Ages. A number of lects are said to be ancestors of Cuman, and it is hard to sort out which ones best represent the ancient tongue.

A good case can be made that the Karaim and Crimean Tatar languages along with Krypchak, Karachay-Balkar, and Kumyk are the best examples of living Cuman languages today. Karaim speakers are the descendants of actual Cuman Kipchak speakers who were converted to Judaism during the reign of the Khazar Khaganate.

Kazakh-Kirghiz

Kazakh-Kirghiz is the name of a single language with two dialects that is generally seen as two separate languages, Kazakh and Kirghiz. Since on a linguistic basis, these lects are best seen as dialects of a single tongue as opposed to separate languages, it is best to characterize them in this way.

The problem is that both Kirghiz and Kazakh are designated as the official languages of two separate nations, Kazakhstan and Kyrgyzstan. Each dialect has its own written version. So for political and social reasons, Kazakh and Kirghiz are regarded as two separate languages, but this characterization has no basis linguistically.

[*Kirghiz*](#) (Lewis et al 2009j) and Kazakh are very closely related, and there is full intelligibility in regular conversations between the two dialects. The vowel system of Kirghiz differs from Kazakh with regard to the presence of certain progressive features such as labialization. Both Kirghiz and Kazakh were regarded as a single language - (Kara) Kirghiz - by Russian linguists in the 1800's (Eker 2015).

The best way to refer to these two lects today is that they are a single language with two dialects, Kirghiz and Kazakh. The best name for it is Kazakh-Kirghiz, following the tradition of Serbo-Croatian where the two major dialects of the language represent the name of the language itself. Karluk is another possible name for this language, but it should be discarded because it also includes Uzbek (Eker 2015). In addition, they have been growing closer to each other recently. Kazakh and Kirghiz only split in recent years. Kirghiz is spoken in Kyrgyzstan.

One view holds that Kirghiz forms [its own separate branch](#) of Turkic (Tekin 2010). But if Kirghiz is really a part of a single language called Kazakh-Kirghiz, then how can the Kirghiz dialect be in one separate branch of Turkic and Kazakh be in another? It makes no sense.

Kirghiz is also close to Altay. The Altay influence is related to the Dzungarian Invasion in 1600-1900 CE, but recently Kirghiz has been strongly influenced by Kazakh.

MI between Turkish and Kirghiz is about the same as with Kazakh - ~15% (Tekin 1978).

The roots of Kirghiz are probably located in the extinct East Kipchak language. The Kirghiz tribe is mentioned already in the Orkhon inscriptions from around 730 CE, and they may have already been speaking East Kipchak by then.

Fergana Valley Kirghiz, Fergana Kipchak, Kuman, or Qomanian is spoken in the south of Kyrgyzstan and in Uzbekistan in the Fergana Valley. This lect is the descendant of the Fergana Kipchak language that went extinct in the late 1920's.

This dialect has a lot of Uzbek influence (Johanson and Ragagnin 2006).

The dialects spoken in Uzbekistan look more Uzbek than Kirghiz. The relationship of this dialect to Khwarezm-Kipčak Uzbek spoken in the same area is not known. Perhaps they are one and the same, or perhaps these are two distinct dialects. The situation remains confused. Intelligibility with either Uzbek or Kirghiz is unknown.

Xinjiang Kirghiz is spoken in Xinjiang, China. This dialect has acquired quite a few Chinese loans (Johanson and Ragagnin 2006). Other than a few Chinese loans in Xinjiang Kirghiz and some Russian loans in Standard Kirghiz, intelligibility with Standard Kirghiz is full (Dwyer 2015).

Kazakh is a dialect of the Kazakh-Kirghiz language spoken in Kazakhstan. Kazakh speakers say that Tatar has 40-50% intelligibility with careful listening. Kazakh may have ~75% intelligibility of Nogay and Uzbek.

Kazakh has near full intelligibility with Kirghiz in daily conversation (Suer 2015). Turkish speakers say their intelligibility of Kazakh is close to zero, at least at first.

Karagash Nogay is an outlying dialect of the Nogay language. It is [spoken](#) in Krasnoyarskiy and Kharabali counties in Astrakhan Province in Russia. They moved there in the late 1700's.

In recent years, Russian linguists have made a case that Karagash Nogay is a separate language and not a dialect of Kazakh or Nogay (Arslanov 1997a).

Nevertheless, we should reject this categorization on intelligibility grounds because Karagash has 98% intelligibility with Kazakh.

Hence, it is therefore a [dialect of Kazakh](#) (Moseley 2010h). The 2010 Russian census reported 34 Karagash speakers.

Khwarezm-Kipčak Uzbek has a [close relationship](#) with Tatar (UCLA Language Materials Project b), but intelligibility with Tatar is probably only 50%.

In its present form, it is spoken in Uzbekistan by nomads in the north-central region in Northern Khorezm in the kishlaks of the Fergana Valley and among the Kuramas of the Tashkent Oasis (Sengupta 2003). It is best seen as a Kazakh dialect (Salminen 2007a). This dialect has a weak status today (Johanson and Ragagnin 2006).

Khwarezm-Kipčak Uzbek is thought to be an ancestor of the extinct Cuman language spoken this region (Howarth 1883). The relationship of Khwarezm-Kipčak Uzbek to the dialects known as Fergana Kipchak, etc. is not known, however they may also be the descendants of Old Cuman based on the alternative names for the lect. We are not even sure they are different; perhaps they are the same dialect. The situation is very confused.

South Kazakhstan Oghuz, traditionally considered part of the Uzbek language, is actually probably similar to Khwarezm-Kipčak Uzbek and hence is a Kazakh dialect. However, originally it was very close to Khwarezm Uzbek Oghuz, which is a dialect of the Northeastern Langar Khorosani Turkic language. In fact, Khwarezm Uzbek Oghuz and South Kazakhstan Oghuz could be seen as the northern and southern parts of a what was formerly single lect.

South Kazakhstan Oghuz was probably also a dialect of this Khorosani Turkic language, but presently it has heavily assimilated to Khwarezm-Kipčak Uzbek to the point where it is part of the same dialect as Khwarezm-Kipčak Uzbek (Eker 2015).

Since the 1920's, it has come under heavy influence of the Kazakh language and the Oghuz elements of this dialect have been steadily decreasing (Ektazarov 2014).

It has also come under the influence of the Karlukized literary Uzbek

language which these people use to write. The case system has been particularly affected by literary Uzbek. There are also influences from Chagatai (Old Uzbek) in the instrumental case (Ektazarov 2014).

Speakers are ethnic Uzbeks. In some towns, there has been a heavy shift to Kazakh. South Kazakhstan Oghuz is seen by speakers as the last remains of an ancient feudal past which is best abandoned, and Kazakh is seen as the language of modernity. The speech of many of the younger generation looks more like either Standard Uzbek or Standard Kazakh than it resembles South Kazakhstan Oghuz. The most fluent speakers are often elderly, in their 70's and 80's (Ektazarov 2014) .

This lect is probably the last remains of the ancient Oghuz language spoken 1,000 years ago. The Old Oghuz traces are mostly seen in the lexicon, which has many old words (Ektazarov 2014).

Sometimes considered endangered, the lect nevertheless has 80,000 speakers (Ektazarov 2014).

Karakalpak is so close to Kazakh that since the Soviet Revolution most Turcologists have referred to it as a [dialect of Kazakh](#) (Tekin 2010). Nevertheless, Karakalpak and Kazakh have separate written forms. Karakalpak has 98% intelligibility with Kazakh; therefore, Karakalpak is a dialect of Kazakh.

Karakalpak has been split from Kazakh for 400 years. However, the situation may be more confused than this as Northeastern Karakalpak is very close to Kazakh (perhaps the 98% intelligibility referred to here) and Southwestern Karakalpak is more similar to Uzbek.

There may be another Karakalpak dialect spoken in Fergana Valley (Menges 1947).

Karachay-Balkar

Karachay-Balkar (Lewis et al 2009f) is spoken by 302,748 people (Tishkov 2009) in parts of the Karachay-Cherkessia and Kabardino-Balkaria Republics (Moseley 2010g) in the Caucasus to the west of Dagestan. Karachay-Balkar is made up of two nearly identical dialects - *Karachay-Baksan-Chegem* and *Balkar*. Karachay-Balkar may have split off from the rest of Kipchak around 850 BCE (Figure 9).

This language has a close relationship with Kumyk (Campbell and King 2013). Karachay-Balkar has some Circassian and Kabardian influences.

Mutual comprehension between Turkish and Karachay-Balkar is very low, perhaps ~15% (Tekin 1978).

Kipchak A is the shorthand for a possible Kipchak dialect spoken by an unnamed group south of Iran, according to reports (Doerfer 1969, p. 13). As far as we know, no Kipchak languages are spoken in Iran at this time. The claim, which dates back to the 1960's, remains unproven as yet. If this dialect exists, it may look something like Karachay-Balkar.

Karaim (Lewis et al 2009g), actually a macrolanguage, is the language of two groups. One is a group of Turkic speakers who were converted to Judaism during the Khazar period and practice a unique form of Judaism called Karaite Judaism. The other group are related non-Jews called Karaylar-Karaites (*Karaylar* means "Karaim" in the Karaim language). This is a syncretic, non-Jewish sect who believes in both Mohammad and Jesus as prophets.

The Karaylar-Karaites are a group of Karaite Jews who left Karaite Judaism in the early 1800's and intermarried with Muslim Crimean Tatars, hence the belief in the prophethood of Mohammad and Jesus.

Linguistically, it is closest to Kumyk and Karachay-Balkar. According to the latest Russian census, only 88 speakers of Karaim are left in the Russian Federation out of a population of 600 Karaim in Moscow, but

other reports put this figure even lower or even say that Karaim is extinct in the RF. Karaim has strong Hebrew influences in its syntax. Turkish intelligibility of Karaim may be anywhere from [~10%](#) (Tekin 1978) to ~65-70% (Eker 2013).

There are actually two varieties of Karaim.

Lithuanian, Northwestern, Trakay or Troki Karaim is now spoken in only in the town of Trakai, Lithuania (Boeschoten 1998) by a small number of elderly people. Lithuanian Karaim has only 50 speakers left (Salminen 2007b) out of a population of 200 Karaim in Trakai and Vilnius. The Karaims arrived in Lithuania in 1397-1398 from Crimea along with the Lithuanian Tatar population, so they have been split from Ukrainian Karaim for over 600 years. This also indicates that Lithuanian Karaim was derived from Crimean Karaim and not Ukrainian Karaim.

Ukrainian, Southwestern or Halich Karaim is spoken now only in the town of Halich, Galicia, in Western Ukraine (Boeschoten 1998). Often thought to be extinct, this language is better seen as moribund as there remain a few, possibly more than a dozen, elderly speakers over age 70 with good knowledge of the language remaining out of a population of 800 Karaim living in Crimea, Kiev and Lvov. However, the language is rarely used in daily life.

Ukrainian Karaim or Southwestern Karaim may be a better descriptor for this language as Halich was only one of two major dialects of Ukrainian Karaim. The other major dialect was called Lutsk.

Lutsk may be extinct, although it was still alive in 1978 (Németh 2011). Lutsk and Halich were probably mutually intelligible.

The differences between Lithuanian Karaim and Ukrainian Karaim are so great that the two communities use Polish or Russian in order to communicate (Johanson and Csató 2010). Intelligibility between the two Karaims is probably ~70% based on glottochronology (Dybo

2006).

Therefore, Ukrainian Karaim and Lithuanian Karaim appear to be separate languages. All speakers of both Karaims are over 60 years old.

Ukrainian Karaim is very close to Nogay ([Tekin 1990](#)). Nevertheless, based on glottochronology, Karaim intelligibility with Tatar, Kumyk, and Nogay may be poor (Dybo 2006).

Crimean Karaim also went [extinct](#) in the early 20th Century. Whether Crimean Karaim was an independent language or a dialect of Halich Karaim is not known. Crimean Karaim was the closest to Turkish of the Karaim lects. Ukrainian Karaim and Crimean Karaim had heavy Slavic influences that extended all the way to morphosyntax.

[Polish Karaim](#) is alleged to be another Karaim dialect, but its existence is dubious (Eker and Kasapoglu 2009). Karaim is said to be extinct in Poland (Salminen 2007b). However, reports say there are 20 older speakers of Karaim remaining in Poland (Boeschoten 1998) out of a population of 150 living in Warsaw, Wroclaw and Gdansk.

After the Bolshevik Revolution, many Ukrainian Karaims fled to Poland. But, many of the original Lithuanian Karaims lived on the border of Lithuania and Poland. The best analysis is that Polish Karaim speakers arrived in the region at the same time as the Lithuanian Karaim and speak Lithuanian Karaim.

Krymchak, the language of the Crimean Jews, is often said to be a dialect of Crimean Tatar.

It is also [very similar](#) to Karaim (Eker and Kasapoglu 2009). However, there are [differences](#) in phonology, morphology and also lexicon between the other two Jewish Turkic languages (Anonymous 2012). Many of the words are different in Krymchak and Karaim. The genesis

of the Krymchak language is somewhat mysterious, but it appears to be a split off Karaim.

The differences between Krymchak and Crimean Tatar are great enough that in 1997, the Russian Academy of Sciences recognized Krymchak as an [independent language](#) (Anonymous 2012). Although intelligibility data between Krymchak and Crimean Tatar is not known, the decision was apparently made because Krymchak is a structurally separate language.

We should follow the lead of the Russians then and recognize Krymchak as a separate language, not a Crimean Tatar dialect (Anonymous 2012).

Most Krymchak speakers were murdered by Nazis in World War 2, and many of the rest went to the US and Israel or assimilated their speech to the Crimean Tatar language. Krymchak is [nearly extinct](#) (Salminen 2007a), with only 29 speakers left (Tishkov 2009), all older than 70 years old, and even they use Krymchak only rarely. The name Krymchak is probably related to the Russian word "Crimchani" which means people who live in Crimea.

Kumyk is [spoken](#) (Lewis et al 2009k) in far northeastern Dagestan by 458,921 people (Tishkov 2009). Kumyk has five dialects - Buinaksk, Khaitag, Khasavyurt, Podgorniy, and Terek - and all are quite divergent; however, they are all said to be [mutually intelligible](#) (Minorities At Risk Project 2006). Kumyk is close to Tatar, Nogay and Crimean Tatar.

By [lexicostatistics](#), Nogay and Kumyk are closer than Tatar and Bashkir, and hence may be part of a single language (Dybo 2006). Kumyk may have split from the rest of Kipchak around 1100 CE (Figure 9).

Kumyk was created by the fusion of Oghur, Oghuz and Kipchak lects piling on top of each other in layers one after the other. It has also

come under considerable influence from Caucasian languages such as Dargwa, Chechen and Avar along with Indo-Iranian languages like Ossetian. Intelligibility of Kumyk and Tatar may be ~80% based on glottochronology.

Kumyk may also have ~80% intelligibility with Kazakh and Nogay (Eker 2013). Intelligibility testing for Kumyk and Nogay would be nice.

Kipchak-Tatar

Tatar

Tatar or Kazan Tatar (Lewis et al 2009) is spoken by 5,347,706 people in Tatarstan (Tishkov 2009). MI between Turkish and Tatar is very low - ~15% (Tekin 1978). However, Kazakh native speakers say that Kazakh has 50% intelligibility of Tatar.

Ural Tatar is spoken by the Ural Tatars who arrived in the region to the north of the Bashkirs migrating from the Volga Region around 1450 CE. Intelligibility with the rest of Tatar is not known. Some say this is a separate language, but until we get better data, we will consider it a dialect of Tatar.

Mishar Tatar is spoken in the region between *Tibir Tatar* and Kazan Tatar. It is sometimes characterized as a separate language, but without intelligibility data, it is not possible to split it, and it must be seen as a dialect of Tatar.

Part of the Mishars' identity in Chuvashia centers around their dialect, and Mishar Tatar's [differences](#) with Kazan Tatar are a big part of Mishar identity (Vovina 2006). However, Kazan Tatars say that Mishar Tatars speak the same language they do. Mishar Tatar is apparently a dialect of Kazan Tatar.

There are reports that Russian linguists believe that Mishar Tatar is a

distinct language separate from Kazan Tatar, but until we get more information, we should not split it off.

Beserman Tatar is a dialect of Kazan Tatar, but it is poorly known.

Mishar Tatar, *Ural Tatar*, *Beserman Tatar* and Kazan Tatar are together as a single group.

Estonian Tatar, Finnish Tatar, Lithuanian Tatar and Polish Tatar are Tatar dialects spoken or formerly [spoken](#) in the Baltic region by the ancestors of Kazan and Mishar Tatars who moved into the area (Eker and Kasapoglu 2009). Polish and Lithuanian Tatar are extinct, but Estonian and Finnish Tatar still have a few speakers.

The Polish Tatars arrived in Poland in the 1390's.

Polish Tatar went extinct around 1700 CE (Eker 2013).

The Lithuanian Tatars are Crimean Tatar and Nogay speakers who arrived in the 1390's CE.

Their language, *Lithuanian Tatar*, went [extinct](#) in 1450 CE (Humphreys and Mits 1991d).

The Estonian Tatars are descendants of Kazan Tatars who arrived in the 1870's. The 2003 census recorded 1,229 speakers of *Estonian Tatar*. Estonian Tatar probably has full intelligibility of Kazan Tatar.

There is a small *Finnish Tatar* community in Finland registered as an official ethnic group of the country. They speak Mishar and often [mix](#) a lot of Finnish in with it (Ståhlberg 2013). The intelligibility of Finnish Tatar with Kazan Tatar is unknown; hence it is not known if it is an independent language or a Tatar dialect.

[*Bashkir*](#) (Lewis et al 2009b) is spoken by 1,379,727 people in Bashkortostan (Tishkov 2009) and [1.6 million more](#) in the Russian Federation (Moseley 2010c). Tatar and Bashkir split around 1550 CE. Bashkir and Tatar are even closer than Kazakh and Kirghiz based on lexicostatistics, and they must have only split in the last 300 years.

They may be better seen as mutually intelligible dialects of a single tongue which are regarded as separate languages for sociopolitical reasons.

In fact, intelligibility between Bashkir and Tatar is over 90%. They differ mainly phonetically while the lexicon and grammar are much the same, although both have standardized written forms that may differ in some ways. Until the 20th Century, both languages used a common written form. Bashkir scholars regard Bashkir and Tatar as a single tongue (Güzel 2015).

Bashkir has a [considerable Hungarian influence](#) due to close contact with the South Mansi language related to Hungarian (Róna-Tas et al 2011). The influence is mainly seen nowadays in Bashkir phonology.

Tatar is close to Nogay, Kumyk and Crimean Tatar. These four languages may have begun splitting up (Figure 9) around 1300 CE. Tatar-Bashkir split off from the rest around 1500 CE.

MI between Turkish and Bashkir is very low, ~10% (Tekin 1978).

Siberian Tatar

Siberian Tatar, consisting of Baraba Tatar, Omsk Tatar and other Tatar dialects, spoken in the [Baraba steppes](#) in the western parts of Novosibirsk Province in Siberia and to the east in the Omsk Region (Salminen 2007a), is a separate language from Kazan Tatar spoken in the Volga Basin to the west (Tekin 2010).

Siberian Tatar is quite archaic, setting it apart from Kazan Tatar, and MI with Kazan Tatar is sometimes said to be impossible. In addition, all three Siberian Tatar dialects, Baraba, Sibir and Tomsk, are quite remote from each other also, and the distance is so great that it is said to often prevent mutual comprehension. In fact, there is said to be a dialect continuum between Siberian Tatar and Kazan Tatar. Generally Siberian Tatar speakers say that they speak a separate language from Kazan Tatar.

The ethnic structure of the Siberian Tatars is complex and includes Kazakh, Uzbek, Kirghiz, Nogay and Oghuz speakers (Güzel 2015). The most significant influences were from Karluk languages. So Siberian Tatar can be seen as a Karlukized Kipchak language, similar to Nogay.

However, speakers of Siberian Tatar say that the dialects can communicate together well, so the situation is confused (Maneshev 2013).

Considering that there is disagreement on the MI of Kazan Tatar and Siberian Tatar, intelligibility testing would be a good idea. Kazan Tatars should be tested to see how much of Siberian Tatar they can understand. All Siberian Tatar speakers have apparently learned to understand Kazan Tatar since it is used in the schools and as a literary language in the Siberian Tatar region (Maneshev 2013).

Siberian Tatar is probably a recent movement of Tatar speakers from the Volga east to Siberia.

Further, new research suggests that Siberian Tatar is more closely related to Siberian languages like Shor and should be placed in the Yenisei Kirghiz group rather than the Kipchak group where it presently resides.

One study examined 40 Siberian Tatar speakers' articulatory phonetics via MRI and found that consonantal pharyngealization exists in Siberian Tatar, similar to Shor, where it was proven to exist 15

years ago (Ryzhikova et al 2012).

In Baraba it is phonemic or even one of the principal features of the consonantal system; in Tomsk it is simply an allophonic variant or even more minimally, in free variation with important sounds; and in Sibir, its status is unclear. In Sibir Tatar, pharyngealization may operate on vowels also in conjunction with vowel harmony, specifically palatal harmony. This would be similar to Tuvan, where a glottal vowel harmony system exists.

The new data lines up with earlier Soviet researchers who suggested on linguistic and anthropometric grounds that the Siberian Tatars and their language were closer to Siberian languages and ethnic groups than to the Kazan Tatars and their language (Ryzhikova et al 2012).

Siberian Tatar is a dialect chain encompassing many different dialects, and there may be up to 300,000 speakers.

A change request was recently turned in to SIL to recognize Siberian Tatar as a separate language. The requester noted that Siberian Tatar was very different structurally from Kazan Tatar. In addition, a well-known Russian linguist supported the claim that Siberian Tatar is a separate language. In light of this recent evidence, it is best to split Siberian Tatar as a separate language.

Tibir, Sibir or Tobol-Irtysh Tatar is an early split from Kipchak that split off [soon after](#) Baraba did around 800 CE. Sibir Tatar is poorly documented.

Some reports say that Sibir Tatar is [close](#) to the Mrass Shor language (Nixon 2005). Sibir Tatar is thought to be midway between Kazan Tatar and Baraba/Tomsk Tatar.

Baraba Tatar split off from the Kipchak around [800 CE](#), so it has had quite a bit of time to separate.

Baraba Tatar is almost extinct and is [spoken](#) on the Baraba Steppes in the western parts of Novosibirsk Province (Moseley 2010b) just north of Kazakhstan 200 miles east of the city of Tomsk. There are 8,000 members of the ethnic group, but only the elderly speak Baraba Tatar.

The 2002 census recorded only eight speakers of Baraba Tatar (Tishkov 2009). The younger ones have all shifted either to Kazan Tatar or Russian.

This dialect has experienced heavy influence from Kazan Tatar in the past century, probably mostly via the educational system.

Recently, a Russian linguist made the case, perhaps on structural grounds, that of the Siberian Tatar lects, at the very least Baraba Tatar was a distinct language and not a Kazan Tatar dialect (Dmitrieva 1997). This study is in line with native speaker and recent linguistic judgment that has suggested that at least some of Siberian Tatar qualifies as more than a Kazan Tatar dialect.

Tomsk or Tom Tatar is spoken to the east of Baraba Tatar and is grouped together with it in a special group. Tomsk Tatar is now nearly extinct.

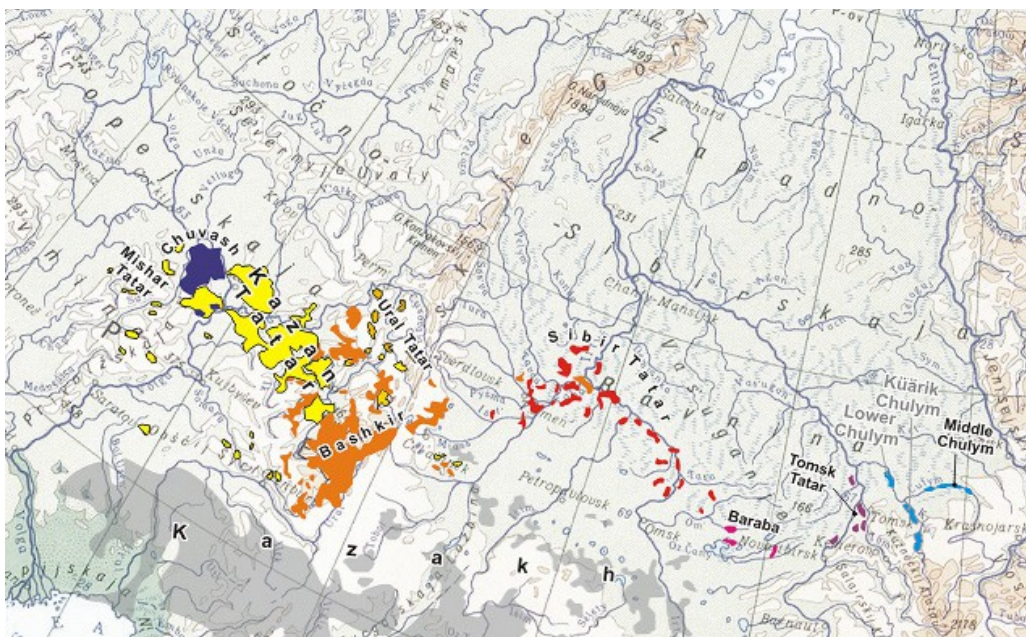


Figure 6: Mishar Tatar (yellow), Kazan Tatar (yellow), Bashkir (orange), Ural Tatar (yellow-orange), Sibir Tatar (red), Baraba Tatar (rose), and Tomsk Tatar (purple). Lower, Kuarik (part of Lower) and Middle Chulym are in blue. From Bruk and Apenchenko 1964.

Nogay and Kazakh are very close but appear to be separate languages. Both are also close to Karakalpak, and Kirghiz is close to but slightly further apart than all three of them. [Nogay](#) (Johanson 2006) has 90,020 speakers in Dagestan and Stavropol (Tishkov 2009). The Nogay people lived with the Crimean Tatar for centuries in Crimea, so there is probably high intelligibility between these two languages (Eker 2013). However, Crimean Tatar speakers in Turkey say that Nogay is [unintelligible](#) to them (Jankowski 2000).

Nogay is also [close](#) to Tatar, Karachay-Balkar and Kumyk. Nogay intelligibility with Tatar, Karachay-Balkar, Kumyk and Kazakh may be ~80%.

On [glottochronology](#), Nogay and Kumyk are closer than Tatar and Bashkir and as close as Kazakh and Karakalpak; hence they may be

dialects of a single tongue (Dybo 2006). Nogay split off from the rest of Kipchak around 1400 CE.

Turkish intelligibility of Nogay may be [very low](#), ~10% (Tekin 1978).

In 1928, the Soviets created two separate literary languages for Nogay, one based on Black (Kara) Nogay and the other based on White (Aq) Nogay. This indicates there may be substantial differences between the two dialects. In 1938, the Kara Nogay literary language was abolished, and Aq Nogay was made the official literary language. Prior to the Revolution, the Nogays used Kazan Tatar and Crimean Tatar as literary languages.

There are three main dialects – Central or Proper Nogay, Qara or Black Nogay, and Aq or White Nogay (Matthews 1951). Differences are often described as slight, but that may not be true.

Nogay Proper or Central Nogay is spoken in the Stavropol region. It has good intelligibility with Qara Nogay and Turkish Nogay.

Qara, Black or Northern Nogay is spoken in Dagestan and like Alabugat Tatar is under heavy Kumyk influence (Boeschoten 1998). Intelligibility with Kumyk is unknown. It has good intelligibility with Nogay Proper and Turkish Nogay.

Aq, White or Western Nogay is spoken in Chechnya and by the Kuban River and its tributaries and in the Mineralnye Vody District in Karachay-Cherkessia. It is under heavy Cherkessian influence (Boeschoten 1998). White Nogay differs markedly from Black and Central Nogay (Matthews 1951). There is a suggestion that Nogay Proper may have difficult intelligibility of White Nogay (Karakoc 2014). This would seem logical due to the significant divergence of White Nogay due to influences from Cherkess. More research is needed to study this matter further.

Alabugat Tatar and Yurt Tatar are two varieties of Nogay that are spoken outside the Nogay region in Southern Russia. They are a group of outlying Nogay dialects spoken in the Astrakhan region of Russia in close proximity to one another. The Nogay ethnic group in that area totals 30,000, but many of them have abandoned their language.

Salminen describes them as [separate languages](#) (Salminen 2007b). Both are almost extinct.

A shift to Kazan Tatar was already underway over 60 years ago . These two dialects are described as “mixed languages”, Alabugat being a Nogay-Kumyk mix and Yurt being a Nogay-Tatar mix (Matthews 1951).

Alabugat Tatar is [spoken](#) on the northwest shore of the Caspian Sea in the village of Severnyy at the station of Ulan-Khol in Lagan’ County in the Kalmyk Republic (Moseley 2010a). The name of the ethnic group is the Utar-Alabugaty. They settled here in the early 1600’s with the Yurts. Like Qara Nogay, this dialect has come under heavy Kumyk influence. The 2010 Russian census reported 1,144 speakers of this dialect.

In recent years, Russian linguists have made a case that Alabugat Tatar is a distinct language as opposed to being a dialect of Kumyk or Nogay (Arslanov 1997a). However, it is uncertain on what grounds they made this judgment. Until we get more information, we should be on the safe side and not split it off yet. In the future, we may be able to split off Alabugat Tatar as a separate language. Intelligibility of this dialect with Kumyk and Nogay Proper is not known.

Yurt or Babayurt Tatar is spoken in Volga (Privolzhskiy), Nariman and Volodarskiy counties in Astrakhan Province, and also in the suburbs of the city of Astrakhan. The Yurts, like the Utar-Alabugaty, came to the area in the early 1600’s. Both groups were fleeing the Kalmyks. Yurt has come under heavy Kazan and Mishar Tatar influence. According to

the 2010 Russian census, there are 31 households where Yurt Tatar is the primary language.

Recently, Russian linguists made the case that Yurt Tatar was a distinct language as opposed to a Kazan Tatar or Nogay dialect (Arslanov 1997c). But it is not known why they made this judgment. Until we get more information, we should not split off this dialect. In the future, we may be able to split off Yurt Tatar as a separate language. Intelligibility data for Yurt Tatar with Nogay Proper and Kazan Tatar is unknown.

There are or were additional outlying Nogay communities in Crimea, Turkey, Romania and Bulgaria.

Dobrujan Nogay is or was spoken in the town of Mihail Kogălniceanu (Karamurat) and the villages of Lumina (Kocali), Valea Dacilor (Hendekkarakuyusu) and Cobadin (Kubadin) in the Dobruja region of Romania and Bulgaria.

Tahsin Gemil, eminent Tatar historian, states that no one speaks Nogay in Dobruja anymore as they have all shifted to Crimean Tatar (Eker 2013). However, a report from 1999 stated that there were still 200 speakers remaining in Romania.

Crimean Nogay is said to be spoken in Crimea in a few villages on the steppes in the far north of the peninsula.

As Turkish Nogays, who speak a Crimean Nogay dialect 200 years removed from Crimea, could communicate with Dagestani Nogays well (Eker 2013), it is highly probable that any remaining Crimean Nogays could likewise communicate well with Dagestani Nogays.

It is probable that many Russian loans have gone into Crimean Nogay in the last 200 years, but Dagestani Nogay has likewise been dramatically affected not just by many Russian loans but by heavy

Russian influence at all levels of the language, including phonetics, so there should be no problems with communication. Further, most Dagestani Nogays are bilingual in Russian, so any Russian loans in Crimean Nogay should not be a problem at any rate (Eker 2013).

The number of Nogay speakers in Crimea is unknown because they are subsumed under Crimean Tatar into one group of 20,000 speakers by SIL. It is not even certain if the Crimean Nogay dialect exists anymore, as, like the Dobrujan Nogays, the Nogays here have been undergoing a process of linguistic assimilation.

As early as 1951, reports stated that the Crimean Nogays were already shifting to Southern Crimean Tatar (Matthews 1951).

[*Turkish Nogay*](#) is spoken in Turkey in small, impoverished villages, often near Turkish Crimean Tatar speakers (Jankowski 2000). Villages are located in southern, western and central Turkey. Over time, 300,000 Nogays have migrated to Turkey, but many have lost their native language and shifted to Turkish.

When Dagestani Nogays came to Turkey recently on a visit, the only language that both had in common was Nogay. Turkish Nogays and Dagestani Nogays were able to communicate very well during the visit, although Turkish Nogays came to Turkey around 1800 from the Crimea, not from Dagestan. Turkish Nogays now feel that Dagestan is their homeland as opposed to Crimea, their true homeland (Eker 2013).

Crimean Tatar

Crimean Tatar, a macrolanguage, has good but not full intelligibility with Turkish.

Crimean Tatar speakers say that Turks cannot understand their language (Dokuzlar 2010). Turkish speakers argue instead that

Crimean Tatar is 100% intelligible with Turkish.

Crimean Tatar does have a close relationship with Turkish due to the Crimeans' close relationship with the Ottoman Empire, whereas the Kipchak languages just to the north such as Bashkir and Tatar lacked this close connection to the Ottomans (Eker 2013).

Intelligibility between Turkish and Crimean Tatar is about 65-70%. However, after a few days' exposure, they can understand each other much better (Eker 2013). Nevertheless, the intelligibility of Crimean Tatar and Turkish may be exaggerated.

Some Crimean Tatar children living in Turkey never learned Crimean Tatar and speak only Turkish. Although Crimean Tatar is spoken around them, these children reportedly [cannot understand it at all](#) (Jankowski 2000). Crimean Tatar and Turkish have separate written forms.

Crimean Tatar is very closely related to but not intelligible with Tatar. Crimean Tatar speakers in Romania were given Tatar textbooks to use, and they could not understand a word in them (Eker 2013). The language is also close to Kumyk and Nogay.

However, Crimean Tatar speakers in Turkey say that Nogay is [unintelligible](#) to them (Jankowski 2000).

Crimean Tatar split off from the rest of Kipchak around 1250 CE (Figure 9). It has experienced considerable Oghuz influence; in fact, some say it is a Kipchak-Oghuz mixed language.

There are [different lects](#) within Crimean Tatar: Northern Crimean Tatar, Central Crimean Tatar, Southern Crimean Tatar, Turkish Crimean Tatar and Dobruja Crimean Tatar (Humphreys and Mits 1991b). The northern dialect (*Northern Crimean Tatar*) has been influenced heavily by Nogay, and the southern dialect (*Southern Crimean Tatar*) has

undergone so much Turkish influence that it is now a dialect of Turkish.

In Crimea where **Northern Crimean Tatar** is spoken, there is difficult intelligibility between the Southern and Northern lects, which makes sense as the former is simply Turkish, and the latter is Crimean Tatar, a different language. Northern Crimean Tatar is spoken by a group called Nogays (not related to the Nogay language spoken to the south in and near the Caucasus) who live on the Crimean Steppe in the north of the peninsula. The base of this language is Kipchak-Nogay.

Central Crimean Tatar is an artificial language created by the Soviets by mixing Northern Crimean Tatar and Southern Crimean Tatar together, but they used Central Crimean Tatar, a descendant of the Cuman language heavily influenced by Oghuz and spoken by the Tat Tatar people of the Crimean Mountains, as a base. This new language is now understood by speakers of both lects.

Central Crimean Tatar is hard to characterize. Although it is intelligible with Southern Crimean Tatar, it does not appear to be fully intelligible with Anatolian Turkish, while Southern Crimean Tatar is. It is probably better seen as a dialect of Northern Crimean Tatar rather than a dialect of Turkish.

It ought to be placed in Kipchak-Tatar along with the rest of Crimean Tatar, although to be completely correct, it would be placed outside of Kipchak-Tatar altogether and put into Kipchak-Cuman, as it is actually a descendant of Cuman. Central Crimean Tatar shows the difficulties, dilemmas and contradictions involved in Turkic classification.

Dobruja Crimean Tatar, the Crimean Tatar spoken in the Dobruja Region of Romania and Bulgaria is said to be a [relatively pure](#) form of Crimean Tatar as Crimean Tatar in Crimea has been heavily influenced by Russian, and Turkish Crimean Tatar has been heavily influenced by Turkish (Jankowski 2000). This is best characterized along with

Northern Crimean Tatar as a Kipchak dialect, but whether it is more Kipchak-Cuman or Kipchak-Nogay is not known.

The genesis of Dobruja Crimean Tatar occurred in 1250 CE when Crimean Tatar speakers came to Dobruja from Crimea, and Turkmen speakers came from Anatolia, and both mixed their languages onto a Slavic-Turkic (Kipchak) mixed base (Csernyei 2014). Since then, Dobruja Tatar has come under [dual foreign pressures, Slavic and Turkic](#) – it is under heavy Bulgarian influence, and the influence of Turkish is also strongly felt (Eker and Kasapoglu 2009). Dobruja Crimean Tatar has full intelligibility of Crimean Tatar in Crimea and hence is a dialect of Crimean Tatar (Eker 2013).

Ukrainian Urum, spoken by 40,000 Urums (Podol'skii 1985) in a dozen villages in Mariupol, Ukraine by Greek Ukrainians, is a dialect of Crimean Tatar. Ukrainian Urum is also close to Karaim, Gagauz and Turkish.

However, a number of [web sources](#) treat Ukrainian Urum as if it is a separate language, and Ukrainian Urums were not happy with education via Crimean Tatar, as they said it was not their language (Anonymous 2013). However, Ukrainian Urum is intelligible with Crimean Tatar (Eker 2015). Based on the preponderance of evidence, at the moment it is best to place Ukrainian Urum as a Crimean Tatar dialect.

Ukrainian Urum is derived from a movement of Greeks out of Crimea to Mariupol in the late 1700's. The story is that Catharine the Great moved them out of the Crimea to Mariupol as part of a political game with the Khanate of Crimea. While in Crimea, they must have [shifted](#) to a Turkic lect that looked a lot like Crimean Tatar and Ukrainian Karaim (Podol'skii 1985).

The Greeks that live near Mariupol speak a Greek lect called Mariupol Greek that may well be a separate language from Greek itself. Ukrainian Urum consists of [12 mutually intelligible dialects](#), one for

each village (Garavets undated).

Ukrainian Urum is in [danger](#) of extinction (Opoudjis 2013), as it is in much worse shape than even Georgian Urum.

Turkish Crimean Tatar is spoken by a group of Crimean Tatars who have been living in Turkey for some time now. Most of them arrived from 1880-1905 as immigrants from both Dobruja and Crimea. Their villages are impoverished, and the language is mostly spoken by people aged 40 and older, especially by the oldest generation.

It is not known whether this language is better characterized as a Kipchak or Oghuz language. Originally it was Kipchak, but it has come under heavy Turkish Oghuz influence. If it is Kipchak, it still remains to be seen if this language is more Kipchak-Cuman or more Kipchak-Nogay.

Language shift to Turkish is ongoing. The group does not want their children to learn Crimean Tatar because they feel it will hinder their Turkish. Turkish Crimean Tatar has undergone heavy changes since it has been in Turkey, and Turkish Crimean Tatar speakers in Turkey say they often have [difficult intelligibility with](#) the Crimean Tatar spoken in the Crimea, mostly due to the huge number of Russian loans that have recently gone into Crimean Tatar and the heavy Turkish influence that Turkish Crimean Tatar has come under (Jankowski 2000).

In addition, Turkish speakers in Turkey, even the children of Turkish Crimean speakers, say they cannot understand Turkish Crimean Tatar at all unless they have learned it, so it is not a Turkish dialect.

Taken all together, Turkish Crimean Tatar is a separate language.

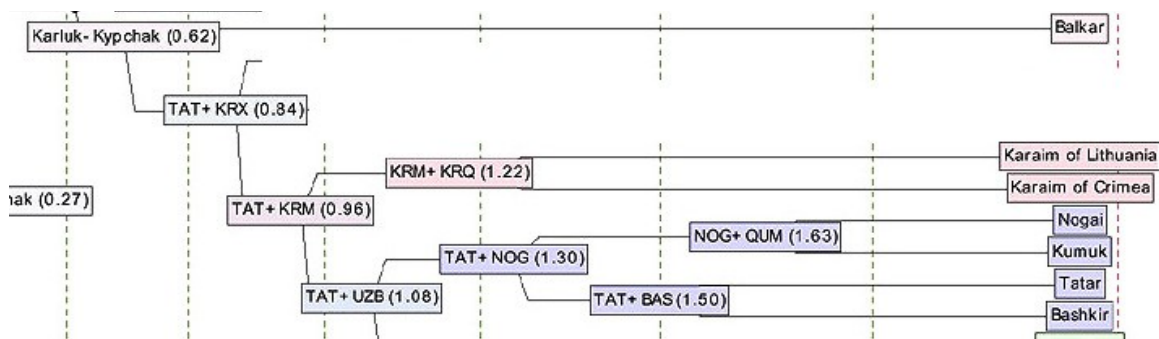


Figure 1: Chart shows the close relationship between Karaim, Nogay, Kumyk, and Tatar-Bashkir. Note that glottochronology shows that Nogay and Kumyk are even closer than Tatar and Bashkir, implying that Nogay and Kumyk may be a single language. From [Dybo 2006](#).

Southern (Oghuz-Seljuk-Karakhanid)

Oghuz-Seljuk

The Southern (Oghuz-Seljuk-Karakhanid) branch is one of the three principal branches of Common Turkic, the other two being Central/Karluk-Kipchak and Yenisei Kirghiz. Most of the languages in this group were part of the Göktürk-Uyghur Empire that ruled from 500-900 CE in Dzungaria, the Gobi Desert, Mongolia and the Tarim Basin. The original languages of this group were Orkhon Old Turkic, Old Uyghur and Karakhanid, which are all extinct.

Facial reconstructions of the Oghuz people indicate that they were originally a Mongoloid people who then became much more Caucasoid after intermixing with the heavily-Caucasoid population in the Near East, generally Iranid tribes and Anatolians.

From 600-750, the Oghuz Confederation called the Tokuz Oghuz was centered around Issyk Kul Lake in far northeastern Kyrgyzstan in the Tian Shan. This area is along the Silk Road. In 744-755, many Turkic tribes fleeing east out of Turkestan in the face of the expanding Uyghur Khaganate may have merged with this group.

Chased out by invaders, they left this area sometime between 750-900 CE and moved down along the Tian Shan west to the Lower Syr Darya River before ending up in the area between the Aral and Caspian Seas around the Urturk Plateau in what is now southwestern Kazakhstan, where they are known from 750-1055 CE as the Oghuz Yabgu State.

By the 1100's, they were conquered by Kipchaks from the Kimak Khaganate. The group then moved down into Iran where they later became the Seljuk Empire, which is the beginning of the Seljuk-Oghuz split. The Oghuz began calling themselves Turkmen and Turcomen starting in the 900's, supplanting the Oghuz appellation. By 1200, the transformation was complete, and the Oghuz no longer existed as an entity.

Oghuz

Turkmen

Turkmen is actually not a single tongue but is more properly seen as a macrolanguage.

Turkmen or Standard Turkmen has fairly low intelligibility with Turkish at around 40%. Turkmen has been strongly affected by Karluk and Kipchak languages. Turkmen is best seen as the living remnant of the earliest Oghuz languages or "Early Oghuz" spoken around Dzungaria in northwestern China. Nowadays this region corresponds to the northern half of Xinjiang Province and is mostly inhabited by Uyghur speakers.

Turkmen has a variety of dialects, including Chovdur, Yomud, Ersarin, Saryn, and Saryq, all of which are quite divergent. The [best available evidence](#) indicates that the Yomud, Ersarin, Saryn and Saryq dialects are all [mutually comprehensible](#) with each other and with Standard Turkmen (Lement 2013).

However, there are reports that when Turkmen was being standardized with the Teke dialect in the 1930's, linguists translated a brochure from Russian into Teke Turkmen. When they took it to Kerki Province, the tribesmen there were [not able to read it](#). This province is now called Lebap Province, and they speak an odd dialect that is a combination of [Turkmen, Uzbek and Tajik](#). This may be the same as the Ersarin dialect. Whether Ersarin is still unintelligible with Teke at present is not known.

The Teke, Yomud, Salir, Gokleng and Alili dialects seem to be less divergent whereas Çovdur, Ersarin and Saryq are further apart (Clark 1998).

Trukhmen is [spoken](#) by 18,000 people (Salminen 2007a) in the Stavropol Region and Astrakhan Province in the Russian Federation and is traditionally said to be a [dialect of Turkmen](#) (Tekin 2010), however, these Turkmen speakers left Turkmenistan around 1700 CE for the Caucasus, and since then, Trukhmen has been under the heavy influence of Nogay and Russian (Eker 2013).

Trukhmen has lost some of its Turkmen features due to prolonged contact with Nogay and other languages in the region. Trukhmen is closer to the Turkmen dialect Çovdur than it is to the official Turkmen language (Demir 2013-2014). Trukhmen intelligibility with Turkmen is ~82% (Demir 2014). This is similar to the intelligibility between other closely related languages such as Czech-Slovak and Spanish-Aragonese/Asturian. Trukhmen then is a separate language.

Seljuk

The Seljuk languages are said to be fully mutually intelligible, but that is not actually the case, as we shall see below.

The Seljuk languages consolidated during the reign of the Seljuk Empire from 1037-1194 CE. Seljuk underwent some minor influences from Kipchak and Karluk, but their impact was not great. With the

defeat of the Seljuks in first East and then West Kharakhanid states by the Kara-Khitans, the Seljuk Empire collapsed. When their leader Ahmad Sinjar died in 1156, the nine atabegs became de facto independent and soon thereafter, effectively collapsed. A while later, only the Sultanate of Rum remained.

The Seljuk languages started splitting up around this time, the Khorasani Seljuqs giving rise to Khorasani Turkic, the Sultanate of Rum giving rise to Ottoman Turkish, the Kermani Seljuks and the Atabeg of Ildeniz probably giving rise to Azeri, the Atabeg of Saljur possibly giving rise to Salchug, the Turcoman Beghlik and the atabegs of Bori and Zangi probably giving rise to Turkmen, and the Kwarzemshahs probably giving rise to the Kwarzem-Oghuz dialect.

The Seljuk lects then have been splitting up for 800 years, nearly as long as Spanish, Portuguese and related lects of Iberia. From 1300-1800, Seljuk languages have come under the influence of Kipchak (via Tatar) and Karluk (via Uzbek).

First of all, it is important to objectively define the boundaries of Western and Southern Seljuk which are frequently blurred. Western Seljuk should be only one group – Ottoman Turkish and related languages – in effect, Turkish, the Gagauzes and Urum. Yet another group is Southern Seljuk, encompassing Azeri and related languages. Many classifications put Turkish and Azeri into the same group, but I feel that this is in error. Azeri, Qashqai, Sonqori, Afshar and Khorasani Turkic are closer to each other than they are to Turkish, Gagauz, etc.

Western Seljuk

Western Seljuk is best seen as a grouping of Anatolian Turkish and related languages that excludes Azeri and related languages which are better put in Southern Seljuk. These are the languages of the Greater Anatolia region that was conquered by the Seljuks and largely converted to Sunni Islam. Most of these people were Turkicized not only culturally but also linguistically.

After the breakup of the Seljuk Empire in 1056, out of nine atabegs, only the Sultan of Rum remained, lasting from 1077- 1307. In the mid-1200's, this reign splintered into small units after the Mongol conquest. By the end of the century, the region began to be transformed into the Ottoman Empire under Osman I, an empire also formed by people who were originally Oghuz.

Turkish

Turkish is of course the most important language in this group, but it is so well known that there is not much to say about it. However, on the question of the Turkish dialects of Eastern Turkey being essentially dialects of North Azeri, this assertion is not correct. True, they are quite close to North Azeri, but they are even closer to Turkish, and speakers cannot communicate easily with North Azeris.

Syrian Turkmen speak a dialect very similar to that spoken by the Iraqi Turkmen, although the dialect spoken by the Syrian Turkmen may be closer to Turkish than the Iraqi Turkmen dialect is. While Iraqi Turkmen is best seen as a dialect of South Azeri, Syrian Turkmen is probably better seen as a dialect of Turkish.

Meshketian Turkish is a dialect of Turkish spoken in Georgia, albeit a divergent one. It is similar to the North Azeri language and the northeastern dialects of Turkish, but it is fully intelligible with Turkish (Eker 2013). They left the Ottoman Empire in 1829. This area of Georgia was conquered by the Ottoman Empire in 1578, and most of this group originated via this conquest. However, a few Turks had migrating into this area since 1000-1200 CE.

Salminen lists *Crimean Turkish* as a separate language (Salminen 2007b). This is reported to be an [outlying dialect of Turkish](#) spoken in a number of villages on the southern shore of Crimea in the Yalta region (Salminen 2007a). There may be 3,000 speakers left, but this may be an exaggeration. Intelligibility with Turkish is not known, nor is their history. Most of this group has probably dissipated by now and

has mingled with Central and Northern Crimean Tatar.

This entry is confusing. It could be a confusion with Southern Crimean Tatar, which is a Turkish dialect. One expert felt that this may be a reference to Yaliboyu Crimean Tatar or the Seashore or Southern dialect of Crimean Tatar, which sounds a lot like Anatolian Turkish (Eker 2013). As Crimean Turkish is often given as another name for Southern Crimean Tatar, this seems to be the best analysis.

Karamanli Turkish was spoken by Greek Christians in Cappadocia, a region in central Turkey. They originally spoke Greek but switched over to Turkish. There were also Slavic (Bulgarian) elements in this speech due to Greeks, Bulgarians and Turks all living in close proximity to each other and mass convergence taking place among the languages. All Karamanli Greeks moved to Greece in the population exchange of 1924. There are still a few speakers of Karamanli Turkish left. Karamanli has full intelligibility with Turkish, over 95% (Eker 2013).

Eastern Anatolian Turkish is one of the most divergent of Turkish dialects. It is often said to be a dialect of North Azeri instead of Turkish. The truth is that this dialect has full intelligibility with Turkish – 95% (Eker 2013). Turkish native speakers say that Eastern Anatolian speakers cannot communicate easily with North Azeris, so it is not a North Azeri dialect.

Southeastern or Kurdistani Turkish is spoken in the center of the Kurdish area of Turkey in the border cities of Mardin, Sirnak, Hakkari and in Bitlis, Siirt and Batman to the north. Turkish speakers say it is hard to understand. The ability of Turkish speakers to understand this dialect depends on their education, language ability and the context of the communication. As a general rule, Turkish has intelligibility of over 90% of Southeastern Turkish (Eker 2014).

Cypriot Turkish is a dialect of Turkish spoken on the island of Cyprus off the coast of Turkey in the Mediterranean. Cypriot has been split

from Turkish proper since 1550 CE, so one would expect some divergence. It has some interesting differences between it and Standard Turkish.

In addition, it has a number of Greek borrowings as the Cypriots live in close proximity to a large Greek community, and many of them speak Greek also. Reports indicate that there are sometimes problems in intelligibility between Cypriot Turkish and Standard Turkish. However, experts state that Standard Turkish has no serious intelligibility problem with Cypriot Turkish (Eker 2014).

Southern Crimean or Yalibolou Tatar is a dialect of Turkish, albeit with some Russian vocabulary (Eker 2013), however the central and northern dialects are different, and intelligibility between them and Turkish is difficult at times. While Southern Crimean Tatar is intelligible with Central Crimean Tatar, it is not fully intelligible with Northern Crimean Tatar.

The fact that Southern Crimean Tatar is fully intelligible with Turkish is responsible for much of the angry confusion surrounding Crimean Tatar and Turkish. Many Turks say that Crimean Tatar is fully intelligible with Turkish, but they are probably referring to the Southern dialect and not Crimean Tatar proper (Central Crimean Tatar) or Northern Crimean Tatar. This dialect is the result of the migration of Anatolian Turks to the Crimea from 1478-1774 when it was still under the shaky control of the Ottomans.

Georgian Urum, spoken by Greeks in K'vemo K'art'li in eastern Georgia, is very close to Turkish because it was derived from a movement of Turkish-speaking Pontic Greeks who left Anatolia around 1800.

Although the base of Georgian Urum is Turkish, and it is closer to Turkish than to any other language, Georgian Urum has many Russian loans.

However, speakers of Eastern Anatolian Turkish in Erzerum, a city in Eastern Turkey, have excellent understanding of Georgian Urum except for some Greek and Russian loans. Nevertheless, initial MI may be hampered by differences in suprasegmentals between the two dialects (Eker 2015).

Georgian Urum is not closely related to Ukrainian Urum.

Georgian Urum has only 1,500 speakers and seems headed for [extinction](#). There is a salvage linguistics project currently going on there via professors from Greece (Skopeteas et al 2011).

Georgian Urum is best seen as a Turkish dialect.

Urum is the name for a cluster of dialects spoken by Turcophone Greeks in Southeast Ukraine, Georgia, and Turkey. Typically represented as a single tongue, the truth is it is not even a language at all, and instead it represents dialects of two different languages.

There are three different types of Urum. Georgian is a dialect of Turkish, Ukrainian Urum is a dialect of Crimean Tatar, and Turkish Urum may be extinct.

Turkish Urum is said to be spoken in Turkey.

However, the evidence suggests that it may be extinct assuming it ever existed in the first place. All Greek Christians left Thrace in 1924, while all Turks left Greece as part of the "population exchange" agreement. For the time being, we will treat Turkish Urum as either a nonexistent or at best an extinct language (Eker 2014).

Gagauz

Gagauz, a macrolanguage, has very high intelligibility with Turkish at

possibly 75-80% based on [glottochronology](#) (Dybo 2006).

Balkan Gagauz Turkish, Danube Turkish or Danubian Turkish is listed as a separate language by [SIL](#) (Lewis et al 2009a), which also claims that Balkan Gagauz is distinct from Gagauz Proper. The Balkan Gagauz Turks live in Bulgaria, far west Turkey, Greece, Macedonia and Romania (Boeschoten 1998), however in Greece and Macedonia, they may no longer speak Balkan Gagauz Turkish.

It is mainly [spoken](#) in the coastal region to the north of and inland from Varna in Bulgaria, which represents the original Gagauz territory.

There are only [400 speakers](#) left in Bulgaria and a few in Romania (Boeschoten 1998), but there are many more in Turkey. In Turkey, this language is often called Thrace Turkish or Rumelian Turkish. Turkish speakers in Turkey say Rumelian Turkish is hard to understand.

Reports from Bulgarians on the Balkan Gagauz Turkish speakers in northern Bulgaria say that Turks find it very hard to understand them, and the Balkan Gagauz find it very hard to understand Anatolian Turkish. There seems to be little doubt that this language is different from Anatolian Turkish.

Khardzali Turkish is another type of Turkish that is spoken in town of Khardzali in the Rhodopean region of southern Bulgaria. Although this is said to be different from Balkan Gagauz Turkish, until we get better information regarding just how different it is from Balkan Gagauz Turkish along with intelligibility information between the two lects, it is best for now to leave it as a dialect of Balkan Gagauz Turkish. More work is needed on this little-known dialect.

Gagauz is spoken by people called the Gagauz, Christians living in the Gagauz Autonomous Region of Moldavia who strangely enough speak a Turkish language with many Slavic loanwords. Gagauz is one of the three official languages of the region, the other two being

Romanian and Russian.

There are accusations that the government is doing little to preserve and promote the Gagauz language here (Anonymous 2014).

Both Gagauz languages are presently being heavily influenced by Turkish.

There is little data available on intelligibility vis a vis Anatolian Turkish and Gagauz, but apparently SIL felt that there enough differences in terms of language structure or intelligibility issues to justify calling it a separate language.

Turcologists living in Turkey say that they can understand over 90% of Gagauz, but they added that they teach Turkish languages and literature, so they are probably not representative of the average Anatolian Turkish speaker (Eker 2015).

The Gagauz arrived in Bessarabia, now Moldova, from their original settlements in Bulgaria about 200 years ago due to persecution by Ottoman Turks. Many Bulgarians moved with them and ended up living around the Gagauz and sometimes intermarrying with them.

But if Balkan Gagauz has low intelligibility with Anatolian Turkish, surely the same language after 200 years in Moldova has not gotten any closer to Turkish. In other words, if Balkan Gagauz has poor intelligibility of Anatolian Turkish, surely Gagauz must also because it is the same language, only different in that it moved from Bulgaria to Moldova 200 years ago.

200 years does not seem to be enough time for a new language to develop over in Moldova, but SIL says that Balkan Gagauz is different from Gagauz, apparently structurally (Lewis et al 2009a). We have no data on MI between Balkan Gagauz and Gagauz. This information is needed if we wish to determine if we have one language here or two.

Southern Seljuk

This group is from the southern part of the Seljuk Empire. This area was conquered by the Seljuks, but the culture and much of the language was left intact. The conquered people were Iranic Muslims, and the Seljuks respected them as fellow Muslims, largely left them alone and did not pressure them to speak Turkic. Persian was even the official language of the Seljuk Empire. Some of these tribes were Turkicized over time, but many retained their Iranian languages. Hence, Turkicization in this region was incomplete.

It was due to the incomplete Turkicization of this area that we see so much Turkic-Iranian mix in these languages such that they are often incorrectly described as Turkic-Iranian mixed languages. At any rate, all Iranian Turkic lects have been very heavily Persianized in lexis and in some cases even in grammar.

The relationship between Azeri and Turkish is often overstated. I do not feel it is correct to put Azeri and Turkish in the same Western Seljuk group.

Azeri is closer to Qashqai, Sonqori, Afshar, and Khorasani Turkic (Turkic languages of Iran) than it is to Turkish and Gagauz. Therefore, placing Azeri in Southern Seljuk and Turkish in Western Seljuk is rational. Southern Seljuk then encompasses Azeri and related dialects and languages such as Qashqai, Sonqori, Afshar and Khorasani Turkic.

The placement of Khorasani Turkic is controversial. It is often placed in a separate group called Eastern Oghuz along with Turkmen. However, since we are splitting Seljuk off from Oghuz, it is best to leave only Turkmen in Oghuz and put all other Oghuzic languages into Seljuk. Khorasani Turkic is in between South Azeri and Turkmen, but it is closer to Turkmen. Hence it is often placed in a group with Turkmen.

However, Khorasani Turkic is closest of all to Afshar which has had tremendous effects on the language. As Afshar is best placed in Southern Seljuk with Azeri-type languages, it would be best to put

Khorasani Turkic there too.

Azeri is a language spoken in two dialects – North Azeri which is spoken in Azerbaijan, and South Azeri which is spoken in Iran.

North Azeri is spoken in Azerbaijan. North Azeri and Turkish are often said to be completely intelligible with each other, but this is not true, although the situation is interesting.

For one thing, many of the [neologisms](#) that went into Turkish following Ataturk's Ozturkche language reforms are not intelligible to North Azeri speakers (Doerfer 1998). In addition, North Azeri retains many of the Persian and Arabic loans that were stripped from the language during the reforms, so North Azeri speakers can nearly understand Ottoman Turkish better than Modern Turkish.

In particular, northeastern Turkish dialects are very close to North Azeri. The Eastern Anatolian dialects of Turkish are closer to North Azeri than to Turkish and are often listed as North Azeri dialects, but the truth is people who reside in the region that say that even these dialects are not fully intelligible with North Azeri.

Turkish-North Azeri MI is good. After a few weeks of close contact, they can communicate fairly well (Eker 2013).

Three intelligibility studies have been conducted on Turkish-North Azeri MI. The first found 49% intelligibility, rather low (Fraenkel 1962). Two other studies were done, one showing 66% and the other showing 92% (Roos et al 2009). The average for all three studies was 69%. The wide range implies that there may be some bilingual learning or other external factor involved which complicates MI calculations.

Turkish-North Azeri intelligibility is overrated, particularly on the Turkish side. Azeris say that Turks and Azeris communicate pretty well

in Baku, but this is only because Azeris adjust their speech to include a lot of Turkish words instead of native Azeri words in order to make themselves more understood. In addition, Azeri speakers who go to Turkey say that they are either not understood or people laugh at them and tell them they speak like a very old Turkish person (in other words, they are speaking an archaic version of Turkish).

This MI of Turkish and North Azeri is changing now due to increased contact. Nowadays due to exposure to Turkish TV, most North Azeri speakers can speak Turkish well, and due to exposure to North Azeri TV, Turks understand a lot more North Azeri than they used to. On the other hand, North Azeris have a lot of motivation to learn Turkish because it is a large, prestigious language, but most Turks have little interest in learning North Azeri, as they do not see it as useful or worth their time.

Azeri is split into the North Azeri and South Azeri dialects. Although the two are completely intelligible, there are large differences in phonology, syntax and loan words.

South Azeri, spoken by [13 million people](#) (Windfuhr 2012) in the northwest of Iran, is a dialect of the Azeri language. South Azeri is best seen as a heavily mixed Turkic-Iranic-Arabic language. They are an Iranian people who lost their Iranian language and shifted to a Turkic dialect.

Intelligibility between Turkish and South Azeri is controversial but seems to be full at over 90% (Eker 2013), however, native speakers say that South and North Azeri are closer to each other than either one is to Turkish (Esla 2014), which implies that while South Azeri is intelligible with both North Azeri and Turkish, it is better seen as a dialect of the Azeri language and not a dialect of Turkish proper.

North and South Azeri have been split for only 200 years, since Iran lost its northern territories in 1812. The MI of South Azeri and North Azeri has been [decreasing](#) in recent decades because when they were

politically separated, South Azeri came under serious Persian influence and adopted many Persian loans, and North Azeri increasingly adopted a more purist official language (Johanson 2011). Nevertheless, South Azeri native speakers say intelligibility between North and South Azeri is 98% (Eshoo 2014).

[*Salčuk or Salchug*](#) (Knüppel 2010) is a poorly known dialect spoken to the south and southwest of Kerman in southeastern Iran. Almost nothing is known of this dialect.

Ethnologue says it was a separate language and claims it is already [extinct](#) (Lewis et al 2009o). The name Salchug was probably derived from the term Seljuk, a well-known term in Turkic linguistics and history (Esla 2014). There are claims that it was probably a dialect of South Azeri.

[*Galūgāh*](#), another Southern Seljuk dialect, is a South Azeri dialect, albeit a divergent one.

It is spoken in Iran on the eastern part of the southern shore of the Caspian Sea (Knüppel 2010). The term Galūgāh may be a geographical term (Esla 2014).

[*Zanjan*](#) is a Turkic dialect spoken in the Iranian province of that name that is transitional between South Azeri and Qashqai (Doerfer 1998). Others imply that it is [a dialect of South Azeri](#) (Grimes 1996). Actually, it is considered to be one of the four main dialects of Azeri, in this case [the southeastern one](#) (Windfuhr 2012). The term Zanjan appears to be a geographical term for the province where the dialect is spoken (Esla 2014).

The people in this region continued to speak an Iranian language until about 300 years ago. After that, many Turkic speakers began migrating to this area from Iranian Azerbaijan and the Ardabil region. They gradually became the majority, and most of the Iranian speakers began to give up their language in favor of the Turkic dialect. Now

Zanjanis are considered a Turkic people, but there are still some villages that still speak an Iranian language. So this dialect is a result of Iranian speakers shifting to Turkic over the last few hundred years.

Teimurtash, Teimuri, Timuri, or Taimour is a dialect spoken in by 7,000 people in Mazandaran, Iran (Grimes 2000a). It is apparently similar to Salchug. Like Salchug, it is said to be a South Azeri dialect (Grimes 1996). This dialect is little known.

The term Teimouri is probably a tribal appellation (Esla 2014). This appears to be a group of Mongols called the Timuri that lives in Khorasan and speaks "Turkish."

They came from Herat in Afghanistan in 1838 and are probably related to the Hazara (Oberling 2008). They should not be confused with speakers of a Mongolian language of the same name also spoken in Iran.

Pichaqchi, Pishagchi, or Bucaqchi is an Azeri-type dialect related to South Azeri, Shahsevan, Aynallu and Afshar. This appears to be another dialect of Azeri.

It is spoken in western Azerbaijan by 1,000 people (Grimes 2000a). This is yet another Oghuzic dialect spoken by a tribe in Iran.

Ethnologue lists Pishagchi (Pichaqchi, Bucaqchi) as a dialect of South Azeri (Lewis et al 2014).

Shahsevan or Shahsavan is an Azeri dialect spoken mostly in Markazi Province, Iran. The Shahsevan say they are not Azeris and that they do not speak Azeri. Instead, they call themselves Shahsevan and say they speak Shahsevan.

Shahsevan may date back only to 1720, when the group presumably

adopted a Turkic tongue due to mass violence displacing settled tribes in the area initiated by a local leader known as Nadir Shah. Prior to that, they spoke some non-Turkic, probably Iranian, dialect. Some believe that they are simply Kurds who adopted a Turkic language.

Ethnologue lists Shaksavani (Shahsevan, Shaksavan) as an Azeri dialect. This dialect has 130,000 speakers. Shahsevan has some structural differences from Azeri, and in some respects it looks more like Qashqai, especially the second person singular verbal declension, and in phonetics, the letter *Q* is pronounced *gh* and not *g* as in Azeri. In Iran, some people say it is a separate language, and others say it is an Azeri dialect.

The reasons for stating it is a language seem to be on sociopolitical grounds as Shahsevan speakers do not live in the Azeri region and do not identify themselves as Azeris or Farsis. They were referred to as frontier nomads in the 1800's, and 50,000 of them still live a nomadic life today.

Like the Aynallus and Afshars, the Shahsevan have roots in the Shahsevan Confederacy. All three - Shahsevan, Aynallu and Afshar - are former Qizilbash Turkmen Confederacies dating back to 1500-1600 CE. Intelligibility between Shahsevan and Azeri is not known but is likely to be good.

Tat is the name of Azeri dialect groups that live near in various places in the Shahsevan speak very similar Azeri dialects. Tats are simply the settled agricultural peoples of the area. Many Tats are former Shahsevan, and many Shahsevan are former Tats who gave up sedentary life to be nomadic.

[*Paradonba, Sharak and Ali-Qurchi*](#) are three dialects spoken in the Iranian villages of Paradonba, Sharak and Ali-Qurchi that are said to be close to Qashqai (Doerfer 1998). However, a better description of them would be to say they are dialects of South Azeri (Dolatkhah 2013).

All eight of the South Azeri dialects above are spoken by Iranians who refer to their dialects as “Turchi,” and local Farsi speakers refer to these dialects as “Torki.” In South Azeri-type dialects, $k > ch$, hence the difference between the Farsi and South Azeri terms for “Turkic” above. This change also applies to most other loans in South Azeri (Esla 2014).

Iraqi Turkmen is better seen as a dialect of South Azeri ([Tekin 1990](#)) than as a dialect of Turkish. Intelligibility with Turkish is around 90%, and while Iraqi Turkmen is similar to both Turkish and South Azeri, it is closer to South Azeri (Eker 2013), and hence is better seen as a dialect of South Azeri than as a dialect of Turkish. Still, the pronunciation is very different from Azeri’s.

It has much Arabic influence and has also been influenced dramatically by Turkish, which the speakers use as a literary language.

It is quite similar to Afshar (Johanson 2001). This dialect has a close relationship with Sonqori in Iran, and there may be a dialect chain between them.

Qashqai, spoken by [570,000 people](#) (Windfuhr 2012) in Nurabad and Firuzabad in Iran, is very closely related to South Azeri, and until recently it was regarded as a South Azeri dialect.

The ancestors of the Qashqai are believed to have originated from near the Caspian Sea (Tehrani 2010). They migrated south to the Zagros Mountains about 500 years ago.

Qashqai has [heavy](#) Persian influences, and Persianization is ongoing (Bulut 2006).

Qashqai has relatively low intelligibility of Turkish and Turkmen - 15% with Turkmen and 30% with Turkish (Dolatkah 2013). Intelligibility

with non-Seljuk languages is not good. In the text below, the expert quoted refers to Qashqai as Kashkay.

"Previously, Kashkay was understood to be a dialect of Azerbaijani (Cafaroğlu & Doerfer 1959: 281). Currently, it is considered a language in its own right. This conclusion is made on the basis of both sociolinguistic data and pure linguistic characteristics. The main argument in this regard is that Kashkay shows a great divergence from other Oghuz languages, particularly Standard Turkish and to a lesser extent Azeri (especially that spoken in the Republic of Azerbaijan).

There is, however, a great degree of interintelligibility (about 70-80%) between Kashkay and Azeri spoken in northwestern Iran. An average Kashkay would understand also other languages of Oghuz branch, such as Turkish and Turkmen, though only to a limited extent. This is mainly because they show significant morphophonological differences." (From Dolatkhah 2012: 32).

Äynallu is typically described as a South Azeri dialect (Lewis et al 2014), but is actually best described as a Qashqai dialect (Dolatkhah 2013). It is spoken by [5,000 people](#) in Iran in the region of Hamadan and Fars Province, especially to the north of Shiraz (Johanson 2001) in the same area as Qashqai. The Aynallu are a former members of the Qizilbash Turkmen Confederacy dating back to 1500-1600. Presumably the dialect dates from this period.

Sonqori (Bulut 2006) is a little known language [spoken](#) in Iran (Knüppel 2010) by 40,000 people in Sonqor (Sunqur), east of Kermānshāh in a large valley set off from the rest of Kurdistan. It is often described as a dialect of South Azeri, but that is not true, as South Azeri native speakers say they cannot understand Sonqori.

Menges early on characterized it with Äynallu and Qashqai as a Southern Seljuk language transitional between South Azeri and Khorasani Turkic (Menges 1951). This implies that it is similar to Afshar, as Afshar is described as transitional between South Azeri and Turkmen. So Sonqori may be similar to Afshar, except it is closer to Qashqai than Afshar is.

But it has [features](#) that set it apart from Afshar (Windfuhr 2012).

Sonqori is also described as [transitional](#) between South Azeri and Qashqai (Doerfer 1998). However, unlike Äynallu, Sonqori is not close to Qashqai (Dolatkhan 2013).

Looking at all of the evidence above, Menges' characterization may be the best one (although it is not close to Qashqai or Khorasani Turkic), and it does appear that Sonqori qualifies as a structurally separate South Seljuk language.

Sonqori is an odd language. In some ways, it shares similarities with the Iraqi Turkmen dialect, and in other ways, there seem to be layers of some more archaic Turkic elements that are hard to characterize.

It has been strongly affected by Indo-Iranian languages, especially Kurdish (speakers are surrounded by Kurdish speakers) and Persian. There is massive Iranian borrowing in the lexicon. It has also borrowed quite a bit of morphosyntax from Iranian. Phonology has also been affected (Bulut 2006).

Some have wondered if Sonqori is Turkic or Iranian.

But the deeply rooted base of the language – the pronouns, locative adverbs and verbs, postpositional morphology, a small base of case inflection, and perhaps most importantly of all, an intact system of verbal paradigms – remains Turkic. In this realm of Sonqori, there has been no borrowing. Nevertheless, much of the case system is being replaced with a borrowed Persian system of prepositional phrases (Bulut 2006).

Sonqori has an uncertain relationship with the very similar nearby Bayat dialect.

[Bayat](#) is typically described as a South Azeri dialect (Lewis et al 2014), but in fact it is a dialect similar to Sonqori spoken in Iran by Kurds near the same region as Sonqori. It is spoken in the isolated

mountain region of Bayadistan east of Hamadan. They are a mixed group. There are also some Bayats in Turkey, but they speak a dialect of Turkish.

The Bayats in Turkey say there are two different Bayat groups in the country, one called Iranian Bayats and another called Mongol Bayats, which may refer to the first and second groups below.

The first group of Bayats is said to have come into the area were one of the 25 original Oghuz tribes. They arrived in Anatolia, Central Asia, Transcaucasia and Iran with the Seljuks, from 1000-1050 (Doerfer 1988). Some moved on from Iran to Kirkuk in Iraq and from there went on to Anatolia. The famous poet of the Ottoman Empire, Fuzûlî, was a Bayat from Iraq who wrote in Azeri.

A second group, related to the Bayids, the third largest of the four main Mongolian clans, is said to have come to Iran via the Mongol invasion of the 1200's.

The present Bayat group in Turkey contains both of these groups.

However, Turkish scholars reject the idea that links the Oghuz Bayats with the Mongol Bayats (Bayeddin 1989).

A different theory notes that the Bayats, a group known as Khorasan Bayats, fought fierce wars with the Mongols, even killing Genghis Khan's son in law. The Mongols later killed many Bayats in revenge. The theory that the Bayats came to Iran with the Mongol army does not necessarily mean that the Bayats were Mongols. Perhaps the Mongols defeated the Khorasan Bayats and then took the conquered group with them to Iran as part of their army (Aydin 1984). This theory helps to explain both their Asiatic appearance and the Mongolic and early Turkic words in the lexicon.

One theory of the meaning of the appellation Bayat is that it is

derived from the Mongolian plural of *bayan* "rich" (Doerfer 1988). Once again this hints at a Mongolian origin of some sort.

Indeed, there is evidence that the Bayats fought for different Mongol armies, including Camoka's and Genghis Khan's. The allegiances of steppe tribes were shaky, and they would fight one group and later unite with their former enemies against another group.

Bayat is interesting as it has Turkic archaisms such as *değın* "squirrel" from Proto-Turkic **dEgıŋ* "squirrel". In Anatolian Turkish, *değın* typically has the meaning of "til, until." The Bayats also are said to have a more Asian appearance than most Oghuz tribes.

There is not as much Iranian influence in this dialect as there is in Sonqori, however, like Sonqori, this dialect has had its entire syntax deeply affected by Iranian. Most of the older Bayat-speaking women are monolinguals (Bulut 2006). Whether this is a dialect of Sonqori or a separate language is not known, as we have no intelligibility data. More data is needed.

Khorasani Turkic

Khorasani Turkic, actually a macrolanguage, is spoken by [400,000 people](#), mostly Turkmen (Johanson and Ragagnin 2006), in the Khorasan region of Iran (Windfuhr 2012), and may have ~40% intelligibility with Turkish. Although Khorasani Turkic is said to be a [dialect of Turkmen](#) (Tekin 1990), it is actually partway between South Azerbaijani and Turkmen and is not a dialect of either; however, it is closer to Turkmen than it is to South Azeri.

The truth is that it is closer to Afshar than to any other Turkic language due to massive Afshar influence on Khorasani Turkic. Khorasani Turkic intelligibility with Afshar is not known, but as Khorasani Turkic is not even intelligible within itself, it cannot be fully intelligible with Afshar. But the language that has influenced Khorasani Turkic more than any other is probably Persian via massive

borrowings.

There are actually three separate languages within Khorasani Turkic.

Khorasani Turkic consists of six lects: Northwest (Bojnurd), North (Quean), Northeast (Gujgt), Langar, South (Soltan-abad), and Southeast (Xarw-e and 'Olya) (Doerfer 1998 p. 275).

These six groups then represent three separate entities: Northwest Khorasani Turkic, North-Northeast-Langar Khorasani Turkic, and South-Southeast Khorasani Turkic. The three are not mutually intelligible, (Doerfer 1998 p. 275) hence they are best seen as separate languages.

Although Doerfer offers the fact that there are transitional dialects between these languages as evidence against their status as languages (Doerfer 1998 p. 275), that seems weak, as quite a few closely related languages have transitional dialects connecting them.

The relationship seems more similar to the Ukrainian-Belarussian-Russian, Czech-Slovak, Bulgarian-Macedonian, Catalan-Aragonese, Asturian-Galician, and Galician-Portuguese relationships - very closely related languages that are not fully intelligible with each other but which have intelligible transitional dialects connecting them.

Northwest Khorasani Turkic is a language consisting of the Bojnurd dialect.

North-Northeast-Langar Khorasani Turkic is a language encompassing the Gujgt, Langar and Quean dialects and possibly Oghuz Uzbek and Dāyi.

Oghuz-Uzbek or Kwarzem-Oghuz is a dialect spoken in that some now place in [its own branch](#) of Oghuz (Tekin 1990). The Kwarzem-Oghuz

dialect has [marginal intelligibility](#) with Uzbek, and many Uzbeks cannot understand it (Cash et al 2012).

Kwarzem-Oghuz is [spoken](#) in the southwest of Uzbekistan in the old Khiva State in the southern Khorezm Province and around Urganch (Cash et al 2012) and in adjacent areas of Karakalpakistan in Bukhara Oblast and in Turkmenistan (Johanson and Ragagnin 2006). It is said that this is an Uzbek dialect that was [influenced](#) by Turkmen (Salminen 2007a), but instead it was probably influenced by Khorasani Turkic, which is close to Turkmen.

It is best seen as an outlying dialect of North-Northeast-Langar Khorasani Turkic spoken in Uzbekistan.

It is separated from the rest of the language by Sariq and Ärsari Turkmen (Doerfer and Hesche 1993). This dialect is even more Oghuz and less Eastern Turkic than Doerfer had assumed (Baldauf 2007b).

Dāyı, Dāyı Turkic or Kārgıl is a Turkic dialect spoken in the far northwest of Afghanistan by possibly 3,000 people. It is said to have elements of both Uzbek and Turkmen in it.

But the best analysis is that this is a dialect of North-Northeast-Langar Khorasani Turkic.

It is particularly close to Kwarzem Oghuz (though with some differences) but is spoken outside of Uzbekistan in Afghanistan. The morphology looks most like Qoraköl Northern Uzbek, a Northern Uzbek dialect. Phonologically it looks a lot like Turkmen, however the morphological similarities with Uzbek cause locals to compare it more to Uzbek than to Turkmen. Dāyı vocabulary is mostly Andkhoy Southern Uzbek, a Southern Uzbek dialect with some Turkmen and Kipchak Uzbek (a Kazakh dialect) influences (Baldauf 2007b).

Locals describe the language as between Uzbek and Turkmen (Baldauf 2007b). This is different from core Khorasani Turkic, which is described as midway between Azeri and Turkmen. However, North-Northeast-Langar Khorasani Turkic may be quite a bit different from the rest of Khorasani Turkic, and instead of being midway between Turkmen and Azeri like the core language, it may be midway between Turkmen and Uzbek.

Dāyī appears to be dying out, and most speakers seem to be elderly. Most speakers have switched over to Andkhoy Southern Uzbek, and the few remaining speakers are bilingual in this Southern Uzbek dialect and sometimes code-switch in it. Many Dāyī also speak Dari Persian. Many of the remaining speakers are semi-speakers who learned Dāyī in childhood but have now forgotten much of their Dāyī. Their Dāyī has become heavily contaminated with Uzbek and can be seen as "Uzbekized." However, a few older speakers resisted this change out of pride for their roots, and their Dāyī has remained intact (Baldauf 2007b).

It is interesting to note that in the home-based schools of the region, the old Turkic literary language Chagatai is used as one of the languages of instruction, often using the literature of Navā'ī or Makhdūmquly's poetry (Baldauf 2007b). Navā'ī was a Uyghur poet from Herat, Afghanistan, who wrote in Classical Chagatai of the 1400's.

There is a long tradition of Afghan Uzbeks using Chagatai as a written literary language (Baldauf 2007b).

The Dāyī reside along the Amu Darya River in the oasis city of Andkhōy and in the neighboring districts of Qaramqōl and Arabgul. Kārgārxāna or Kārgilxāna is a quarter of Andkhōy named after the most important subgroup of the Dāyī, the Kārgillā. The majority of the Dāyī community in the city is located in the southern portion of the center of the city. Other Dayi speakers reside in Bāyibustān, Quryān, Lāfi, around the Bālāisār Fortress in Tavāči in a locale called Supijānāyā Kōčasi, in Tōqquzdaray and in the southeastern village of

Ältiböläk. However, Däy speakers in Ältiböläk speak Turkmen, not Däy (Baldauf 2007b).

The Däy are subdivided into a number of subtribes, including the Gärxur; the Kārgillä/Kā:rgär/Kārgällä/Kā:rgıl/Kārgil (the most important subgroup); the Nisāylä/Nisailä, of which the Šāyällä are a subgroup; the Qaradaha, which may not exist; the Sābätfurüş; the Sārqamış/Sāriqamış; the Töqsawāyli/Töxsawāi/Töxtawāy; and the Töqumöyri (Baldauf 2007b). Most of these groups may no longer speak Däy. Gärxur and Kārgillä may be the same group as may the Töqsawāyli and Töqumöyri.

A number of groups around the Andkhöy Oasis, including the Arab, the Aymāq, the Dä:yäkči or Dähyäkči, the Kiyikči, the Jat, the Jögi, the Löli, the Qızılbaş, the Urgänji and the Šıx are said to speak unknown Turkic lects, at least as second languages (Baldauf 2007b).

Judging by their name, the the Dä:yäkči or Dähyäkči may well speak Däy.

The Aimaq and the Qizilbash are well-known ethnic groups in Afghanistan encompassing up to hundreds of thousands of members.

Aymāq refers to the Aimaq, a group of nomadic pastoralists who live in western Afghanistan and the Khorasan region of Iran with a few in Turkmenistan, but they speak Aimaq Persian, a dialect of Persian, not Turkic. Aymāq is a Mongolian word meaning "tribe" or "grazing territory."

The Qızılbaş are probably related somehow to the old Qizilbash Turkmen Confederacy dating from the 1500's. Aynallu Qashqai and Afshar speakers are both part of this confederacy. This suggests that the Qızılbaş may have some relationship to Aynallu and/or Afshar speakers.

Arab is probably short for Arabgul, one of the districts that border on the city of Andkhöy where a number of Dāyɪ speakers reside.

The Urgänji live in a neighborhood in Andkhöy called Urganjikhāna and are said to speak Dāyɪ, but further investigation reveals that they speak a "special language" like the Kiyikči (Baldauf 2007b).

Dāyɪ speakers say that the Kiyikči do not speak Dāyɪ but instead speak a "special language" like the Urgänji (Baldauf 2007b).

Ingeborg Baldauf, the pre-eminent scholar of Dāyɪ searched for the Kiyikči language in this region from 1978-2007. She was always told that the language existed, but for over 20 years, she was not been able to find anyone who speaks it (Baldauf 2007a).

The nature of the Urgänji and Kiyikči "special languages" is not known.

Urgänji was a term formerly used by residents of Bukhara to refer to the Kwarzem Oghuz. The group is named after the town of Urgench (Khalid 1998).

Löli was a general term formerly used for the Gypsies of Central Asia who were defined more by profession than by ethnicity (Khalid 1998).

The term Dähyäkči or Dä:yäkči is probably related to the Uzbek word *dahyakchi* "harvest tax collector, sharecropper under the *dahak* system" (Dirks 2005). Like Löli, this term seems to refer more to a profession than a specific ethnic group.

Nothing is known of the Jat, the Jögi, and the Šix.

The Turkic lects that any of these groups speak are not known, and the entire situation is very poorly researched. In the absence of data

to the contrary, a good provisional hypothesis is that these tribes may speak some sort of Khorasani Turkic lects, perhaps similar to Dāyī.

It is not known for sure where the Dāyī originated, but there are a number of theories (Baldauf 2007b).

Some mention a tribe of Teke Turkmen speakers called the Karagel' that existed in the 1800's Dāyī (Baldauf 2007b). The name closely resembles the name of the largest group of Dāyī, the Kārgillā.

Some Dāyī cite the oral traditions of their elders, who say that the Dāyī migrated from Iran to Afghanistan 200-300 years ago. A local ethnographer stated that the Dāyī or Dā:ī:lar as he referred to them, said that their homeland was a place called Sarīyāyāč in Iran (Baldauf 2007b).

The historical document *Fihris at-Tavārīkh* written by Rizā Qulī Khān, mentions that the leader of the region at the time, Shāh 'Abbās, imported a group of Afshar speakers into Andkhoy. The Andkhoy ruling class was partly Afshar as recently at the early 1800's. Sufijān Āyā was an important Dāyī scholar from Andkhoy 60-70 years ago, he claimed that the Dāyī descended from the Afshār or Awshār, a Turkmen tribe (Baldauf 2007b).

The theories about the Dāyī being related to Afshar speakers are uncertain because Dāyī does not look much like Kabuli Afshar, the Afshar dialect spoken in Afghanistan (Doerfer and Hesche 1989). Nevertheless, Khorasani Turkic, of which Dāyī is a dialect, has been heavily influenced by the Afshar language, so a connection with Afshar is possible on linguistic grounds. The Afshar theory is the only one that has at least some supporting linguistic evidence, as the linguistic connection between Dāyī and the Turkmen language is poor.

Some locals state that the Dāyī originate from the Qara, a Turkmen grouping (Baldauf 2007b).

The name Qara once again resembles Kārgıl, another name for the Dāyı dialect.

The sources said that after they moved to Andkhoy, the Dāyı tongue of the Qara underwent major structural changes so that now the only remains of Turkmen influence are in some of the phonology. There may be something to this theory, as the Qara grazing grounds in the 1800's were to the west of Andkhoy (Baldauf 2007b).

Another theory is that the proposed that the original name of the Kārgillä or Dāyı was *Karkililär*, "people from Karki." Karki is a town in Turkmenistan located 75 miles north of Andkhoy on the Amu Darya River. Andkhoy is home to many recent immigrants, but most of these people are aware of their status as muhajir and the Dāyı do not refer to themselves this way. If the Dāyı indeed came from Karki, they did not come in the last 100 years (Baldauf 2007b). But they may have come from Karki 200-300 years ago.

Both of the theories of a Turkmen origin for the Dāyı suffer from a lack of linguistic evidence (Baldauf 2007b) as the connection between Dāyı and the Turkmen language is not good.

Southern Khorasani Turkic is a language encompassing the Southern Soltan-abad and the Southeastern Xarw-e and 'Olya dialects.

Afshar, Afsar or Afsari is typically said to be a dialect of South Azeri (Lewis et al 2014), but this is not the case.

It is spoken by 600,000 people in Turkey, Syria, and especially Iran - where it is spoken by [290,000](#) people - and Afghanistan (Grimes 1996). The speakers in Iran are in the Beyadistan and Hamadan regions. Like Aynallu, the Afshar are a former Qizilbash Turkmen Confederacy dating back to the 1500's.

There are [two sets](#) of Afshar dialects in Iran (Knüppel 2010). The Afshar language is spoken differently in different parts of Iran.

One group of speakers, the larger one, is in the Qazvin area and northeast of Tehran, including the group speaking the *Solaymānābād* dialect of Afshar southwest of Hamadān. This group is [close](#) to South Azeri.

The second and smaller group speaking the *Pugerd* and *Āštiān* dialects of Afshar is more closely related to Qashqai. These dialects are spoken to the north of the Khalaj-speaking area.

It is this dialect that is [the source](#) of the dialect spoken by the Afghan Afsharis (Windfuhr 2012).

Another dialect commonly referred to in the literature is [Kabuli Afshar](#) (Grimes 1996), [spoken](#) in Kabul, Afghanistan (Knüppel 2010). The Kabul dialect is spoken in what is described as the homeland of Afshars, the place where they came to Iran from.

All three of these groups are described as [transitional](#) between South Azeri and Qashqai (Doerfer 1998).

Afshars in Turkey are said to speak Anatolian Turkish. Perhaps they dropped their language upon the move to Turkey.

Afshar has experienced very heavy borrowing from Persian, especially Dari Persian in Afghanistan where the language originated, such that some have the impression that it is a mixed Dari-South Azeri language. However, it is better to see Afshar as primarily a Turkic language than as a mixed language.

Afshar is not as close to South Azeri as Qashqai is. The specialist view for now is summarized by Gerhard Doerfer who says Afshar is [not a](#)

[South Azeri dialect](#) (Doerfer 2011) and instead has a close relationship with Qashqai. Intelligibility with Qashqai is not known.

However, the language that is closest of all to Afshar is Khorasani Turkic, which has experienced huge Afshar influence.

South Azeri speakers state that Afshar is not an Azeri dialect. Other South Azeris say that they can understand Afshar well, and some Afshars agree that South Azeris can understand them (Azertos 2006). Yet South Azeris often say the same thing about Qashqai, and Qashqai has only ~75% intelligibility with South Azeri. However, people often interpret 70-80% intelligibility as full intelligibility. For instance, many Turks say they have full intelligibility of Azeri (~69% average over two studies).

At the moment, the best way to view Afshar is as a separate language with a close relationship to Khorasani Turkic. Intelligibility testing for this language would be very helpful.

Beriberi is a [hypothesized Turkic language](#) (Knüppel 2010) that may be spoken in a few villages in Khorasan, Iran. Its possible existence was mentioned by Doerfer (Doerfer 1969, pp. 15-16). In fact, Beriberi is not even a Turkic language at all. The truth is that a group called [the Hazara-Berberi](#) (Bethany World Prayer Center 1997) does indeed live in that part of Iran. This is a large nomadic group of 67,000 people that immigrated to Iran from Afghanistan. They speak [Hazangari](#) (Bethany World Prayer Center 1997), an Indo-Iranian language that is very closely related to Persian.

TABLE 1
Oghuz Languages

1. Western Oghuz		
	1.1. Western Oghuz (that is, "Turkish")	
	1.2 Central Oghuz	
		1.2.1. Azerbaijani, or, to be more exact, southern Azerbaijani dialects
		1.2.2. Dialect of Galugāh (eastern coast of the Caspian Sea)
		1.2.3. Salčuk (south and southwest of Kermān)
	1.2 Southern Oghuz	
		1.3.1. Afšār (western and southwestern Iran, and in Kabul)
		1.3.2. Qašqā'i (the provinces of Hamadan and Fārs, especially in the region north of Shiraz)
		1.3.3. Āynallu (the provinces of Hamadan and Fārs, especially in the region north of Shiraz)
		1.3.4. Sonqori (in Sonqor/Sunqur, east of Kermānšāh)
2. Eastern Oghuz		
	2.1 Northeastern Oghuz	
		2.1.1. Khorasani (Khorasan-Turkish, northeastern Iran, province of Khorasan)
	2.2 Northwestern Oghuz	
		2.2.1. Turkmen (northeastern Iran, province of Khorasan)

Figure 2: Table showing the Oghuz languages in a popular current version of the family.

The Internal Classification of Turkic

What follows is my attempt at an internal classification of the Turkic languages.

I moved Khalaj, Yugur-Salar and their parents Kharakhanid and Old Uyghur subsumed under a group called Uyghur-Kharakhanid into a separate level of Turkic on the same level as Common Turkic, that is, Uyghur-Karakhanid is one node and Common Turkic is another node. I

did this to show what I feel is the profound divergence of the Khalaj and Yugur-Salar languages.

Originally there was Bulgaro-Turkic. Oghuzic split off first, of which only Chuvash survives.

A unit I call Orkhon-Kharakhanid-Uyghur broke off from this. The first breakaways from Turkic were two languages - Old Orkhon Turkic known from the Orkhon script carved into stone - and Uyghur-Karakhanid.

Several languages then broke off from Uyghur-Karakhanid, including Old Uyghur, North Kharakhanid, and South Kharakhanid. Of this group of four languages, only Khalaj and Yuguric remain.

Yakutic also deserves its own branch, as it is quite archaic and is not properly linked with any group, even the Siberian group (more of a geographic grouping than anything else).

Some would place Khalaj into a large section of Southern Turkic on a par with Oghuz-Seljuk. This is because this archaic language has spent so long in Iran that a great deal of Oghuz-Seljuk influence has gone into it via borrowings as late influences.

But are these superficial late borrowings what is really striking about Khalaj? Of course not. What is striking about it is its deep archaic nature, and the best way to portray that is via a very deep high-level ranking just below Chuvash. Putting Khalaj in Oghuz is like putting English in Romance due to all the French and Latin borrowings.

If one is to split off Khalaj as Doerfer has done in his latest classification, one really must split off Yakutic also. Yakutic has not been convincingly shown to be related to any other Turkic group, though attempts have been made to connect it with Yenisei Kirghiz, especially Tuvan. These attempts are not yet wholly convincing.

Yakutic is nearly as far away from the rest of Turkic as Khalaj is. In order to represent its profound differences, it is best to split it off as a node from Macro-Turkic.

Beyond Khalaj, Yuguric, Yakutic and Chuvash, I place all of the rest of Turkic into a large grouping that I call Turkic Proper or Common Turkic. This is the group that we usually think of when we think of Turkic, and this group in general can be shown to be a relatively tight-knit grouping.

How to Read the Table

Any entry following a large capital letter is a full language (A., B., C., D., etc.). Entries above those with no letters in front of them are families and subfamilies. Entries following small letters are dialects (a., b., c., etc.). Entries following two small letters (aa., ab., ac., etc.) are subdialects.

* Indicates an extinct language.

? Indicates that there is some sort of a question about the lect.

The questions fall into two areas.

The first area refers to doubt as to whether the lect even exists at all anymore.

The second area refers to known existing lects for which the status about whether they are full languages or dialects of another tongue is still very much up in the air. These lects generally have a close relationship with a known full language which points towards a dialectal classification, but on the other hand, there is also evidence that they may be fairly divergent, which indicates that they could be full languages.

All lects with a ? after them are urgently in need for further work to clear up questions about either their existence or the language vs. dialect problem.

Bulgaro-Turkic

Volga Bulgaric
A. Chuvash

Macro-Turkic

Orkhon-Kharakhanid

Orkhon Old Turkic

Old Uyghur-Karakhanid
A. Old Uyghur*

Yuguric
B. Yugur (Western Yugur, Yughur)
C. Western Salar
D. Eastern Salar

Karakhanid
D. North Karakhanid*
E. South Karakhanid*

Khalaj
F. Khalaj
G. Western Khalaj

Yakutic

Yakut
A. Yakut
B. Dolgan
a. Northwestern Yakut

Common Turkic (Turkic Proper)

Central (Yenisei Kirghiz)

Altay

Northern Altay
A. Kumandin
B. Chelkan (Kuu-Kizhi)
C. Tuba (Tubalar, Tuu-Kizhi)

Southern Altay

D. Southern Altay Proper (Altay Kizhi, Oirot Altay)

E. Teleut (Telengut)

a. Telengit

Khakassian

Khakas

A. Khakas

a. Kamas

b. Shor Khakas

c. Koibal

d. Yarin

e. Sagai

f. Xaas (Kacha, Kachin)

g. Beltir

B. Xyzyl

C. Fuyü Gïrgïs

Shor

D. Southern Shor (Kondoma Shor)

E. Northern Shor (Mrass Shor)

Ös (Chulym)

F. Ös

a. Upper Chulym (Tutal Chulym)

b. Middle Chulym (Melet Chulym)

ba. Yachin

bb. Kumysh

bc. Shuy

bd. Kamlar

G. Lower Chulym

a. Küärik

Tuvan

Steppe Tuvan

A. Tuvan

a. Altai Tuvan (Altai Sayan, Tuwa, Menggu)

b. Dzungarian Tuvan (Jungar Tuvan)

ba. Tsengel Tuvan

**bb. Monchak (Hovd, Kovd, Monjak,
Mondzhak, Kök-Monchak)**

**B. Tuha (Tuhalar, Toha, Uighur Uriangkhai,
Uriankhai)**

Taiga Tuvan

C. Tofa (Tofalar)

E. Soyot

F. Dukha (Tsaatan)

G. Todzhin (Todzhu, Northeastern Tuvan)

H. Southeast Tuvan (Kyzyl Tuvan)?

Kipchak-Karluk

Karluk

- A. Kazakh-Kirghiz
 - a. Kirghiz
 - aa. Xinjiang Kirghiz
 - ab. Fergana Valley Kirghiz?
 - b. Kazakh
 - ba. Karagash Nogay
 - bc. Khwarezm-Kipčak Uzbek
 - bd. South Kazakhstan Oghuz
 - c. Karakalpak
 - ca. Northeastern
 - cb. Southwestern
 - cc. Fergana Valley?

Chagatai

- A. Uyghur (New Uyghur)
 - a. Ili Turki
- B. Lop Nur
- C. Aynu

Uzbek

- B. Northern Uzbek
- C. Southern Uzbek

Kipchak

Cuman

Karachay-Balkar

- A. Karachay-Balkar
 - a. Karachay
 - b. Balkar
 - c. Kipchak A?

B. Kumyk

Crimean Tatar

- C. Crimean Tatar
 - a. Northern Crimean Tatar
 - b. Central Crimean Tatar
 - c. Dobruja Crimean Tatar
 - d. Ukrainian Urum

D. Turkish Crimean Tatar

Karaim

- F. Ukrainian Karaim**
- G. Lithuanian Karaim**
- H. Crimean Karaim?**
- I. Krymchak**

Kipchak-Tatar

- A. Tatar (Kazan Tatar)**
 - a. Ural Tatar**
 - b. Beserman Tatar**
 - d. Polish Tatar**
 - d. Lithuanian Tatar**
 - e. Estonian Tatar**
 - f. Bashkir**
- B. Mishar Tatar?**
- C. Finnish Tatar**
- D. Siberian Tatar**
 - a. Tomsk Tatar (Tom Tatar)**
 - b. Tibir Tatar (Sibir Tatar, Tobol-Irtysh Tatar)**
 - c. Baraba Tatar**
- E. Nogay**
 - a. Turkish Nogay**
 - b. Dobrujan Nogay**
 - c. Crimean Nogay**
 - d. Yurt Tatar?**
 - e. Alabugat Tatar?**

Southern (Oghuz-Seljuk)

Oghuz

Turkmen

- A. Turkmen**
 - a. Yomud**
 - b. Ersarin**
 - c. Saryn**
 - d. Saryq**
 - e. Teke**
 - f. Salir**
 - g. Gokleng**
 - h. Alili**
- B. Trukhmen**
 - a. Chovdur**

Seljuk

Western Seljuk

- A. Turkish
 - a. Meshketian Turkish
 - b. Karamanli Turkish
 - c. Eastern Anatolian Turkish
 - d. Southeastern Turkish
 - e. Southern Crimean Tatar (Yaliboyu Crimean Tatar)
 - f. Syrian Turkmen
 - g. Georgian Urum
 - h. Cypriot Turkish

Gagauz

- C. Gagauz Proper
- D. Balkan Gagauz Turkish

Southern Seljuk

- A. Qashqai
 - a. Āynallu
- B. Azeri
 - a. North Azeri
 - b. South Azeri
 - ba. Salčuk (Salchug)
 - bb. Galūgāh
 - bc. Zanjan
 - bd. Teimurtash (Teimuri, Timuri, or Taimour)
 - be. Syrian Turkmen
 - bf. Paradonba
 - bg. Sharak
 - bh. Ali-Qurchi
 - bi. Shahsavan
 - bj. Tat
 - bk. Pichaqchi
- C. Sonqori
 - a. Bayat

Khorasani Turkic

- D. Northwest Khorasani Turkic (Bojnurd)
- E. North-Northeast-Langar Khorasani Turkic
 - a. Gujgt
 - b. Langar
 - c. Quean
 - d. Oghuz-Uzbek Northern Uzbek (Kwarzem-Oghuz Northern Uzbek)?
 - e. Dayı?
- F. Southern Khorasani Turkic
 - a. Southern
 - aa. Soltan-abad
 - b. Southeastern
 - ba. Xarw-e

bb. 'Olya

G. Afshar

* Extinct

? Existence or language status uncertain.

Thoughts on Pan-Turkic Intelligibility

The intelligibility of Turkish and the Central Asian Turkic languages like Uzbek, Kazakh, Kyrgyz and Turkmen is much exaggerated.

Speakers of these languages who went to study in Turkey in recent years said they had problems with the Turkish language. After the fall of the USSR, many people from these newly independent nations visited Turkey, but they had to bring interpreters with them to communicate with Turks. It's true that Turkish TV is not much watched in the Central Asian Turkic nations, but probably the main reason for that is because Central Asian Turkic speakers can't understand it. They can't even understand the simplified Turkish used in these broadcasts.

Ager, Simon (1998-2013). "[Shor \(Шор тили\)](#)." *Omniglot*.

Anderson, Gregory D. S. (1998). "[Historical Aspects of Yakut \(Saxa\) Phonology](#)," in *Turkic Languages 2* (2): 1-32. Wiesbaden: Harrassowitz Verlag.

Anderson, Gregory D. S. (2004). [The Languages of Central Siberia: introduction and overview](#). Leipzig, Germany: Max Planck Institute for Evolutionary Anthropology.

Anderson, Gregory D. S. (2005). "Indigenous sources of variation in Abakan Khakas," in [Language contact in South Central Siberia](#). Wiesbaden, Germany: Otto Harrassowitz Verlag.

Anderson, Gregory (May 21, 2012). "[NG explorers help record Xyzyl language](#)." National Geographic Society.

Anderson, Gregory D. S. Linguist at Living Tongues Institute for Endangered Languages. Personal communication. December 2013.

Anderson, Gregory D. S. and Harrison, David K. (2003). "['Natural' and obsolescent change in Tofa](#)." Salem, Oregon: Living Tongues Institute for Endangered Languages.

Anderson, Gregory D. S. and Harrison, David K. (2004). "[Shaman and Bear, Siberian prehistory in two Middle Chulym texts](#)," in Vajda, Edward J., editor. *Languages and prehistory of Central Siberia*. Amsterdam: John Benjamins Publishing.

Anderson, Gregory D. S. and Harrison, David K. (2006). "[Ös Tili \(Middle and Upper Chulym dialects\): Towards a comprehensive documentation](#)." Salem, Oregon: Living Tongues Institute for Endangered Languages.

Anderson, Greg and Harrison, David K. (July 7, 2008). "[Enduring Voices field expedition to Siberia \(Chulym River region\)](#)." Salem, Oregon: Living Tongues Institute for Endangered Languages.

Anonymous (December 21, 2012). "[Памятники письменности крымчаков \(Krimchak inscriptions\)](#)." *Why Save Languages?* Tacoma, Washington and Simferopol, Ukraine.

Anonymous (March 9, 2013). "[Crimean Tatar revival](#)." *Why Save Languages?* Tacoma, Washington and Simferopol, Ukraine.

Anonymous (August 14, 2014). "[Gagauz authorities say Gagauz language is not needed](#)." *Turkists*.

Arslanov L. Sh. (1997a). "Alabugatskikh tatar (nogaitsev) iazyk," in *Iazyki mira - Tyurkskie iazyki*, p. 172-187. Moscow: Indrik.

Arslanov L. Sh. (1997b). "Astrakhanskikh nogaitsev-karagashei iazyk," in *Iazyki mira - Tyurkskie iazyki*, p 187-194. Moscow: Indrik.

Arslanov L. Sh. (1997c). "Yurtovskikh tatar (astrakhanskikh nogaitsev) iazyk," in *Iazyki mira - Tyurkskie iazyki*, p. 506-513. Moscow: Indrik.

Atwood, Christopher P. (2004). "[Tuvans](#)," in *Encyclopedia of Mongolia and the Mongol Empire. Modern World History Online*. New York: Facts On File, Inc.

Aydin, Mehmet (1984). *Bayat-Bayat boyu ve Oguzlarin tarihi*. Ankara: Hatiboglu Yayınevi.

Azertos (January 21, 2006). "[Afshar/Qaskay Turks](#)." MEIC - Middle East countries - Iran. *Middle East Information Center*.

Baldauf, Ingeborg. (2007a). "Sprache als Baustein für eine neue Freiheit," p. 33, in *Forschung Spezial: Geisteswissenschaften*. Bonn: Deutsche Forschungsgemeinschaft (German Research Foundation, DFG)

Baldauf, Ingeborg. (July 2007 or 2007b). The Dāyır~Kārgıl of Andkhoy: Language, History and Typical Professions: Discourses on Local Identity. ASIEN 104, pp. 135-152.

Bavuu-Syuryun, Mira (2011). "[Sociolinguistic notes on Tuvinians China](#)." *The New Research of Tuva* 1.

Bavuu-Syuryun, Mira and Aelita Salchak (October 21, 2012). "[Philologists work on teaching Tuvan language issues in Mongolia](#)." *The New Research of Tuva* 4.

Beaumont (July 17, 2013). "[Chulym - the unwritten language of the Turkics](#)." Bangkok: *Freelang Magazine*.

Berta, Arpad (1998). "[Turkic languages of Iran](#)," in Johanson, Lars and Éva Á Csató, editors. *The Turkic languages*. London: Routledge.

Bethany World Prayer Center (1997). "[Prayer profile - The Hazara-Berberi of Iran](#)." Colorado Springs, Colorado.

Bitkeeva, A. N. (2005) "[The Kumandin language](#)," in Funk, D. A., V. Y. Gusev and A. N. Bitkeeva, editors. *Endangered languages of indigenous peoples of Siberia*. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Boeschoten, Hendrik (1998). "[Speakers of Turkic Languages](#)," Johanson, Lars and Éva Á Csató, editors. *The Turkic languages*. London: Routledge.

Bruk, S. I. and V. S. Apenchenko (1964). *Atlas narodov mira* [Atlas of the peoples of the world]. Moscow.

Bulut, Christiane (2005). "[Iranian influences in Songor Turkic](#)," in Csató, Éva Ágnes; Bo Isaksson, and Carina Jahani, editors. *Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic*, 241-271. New York: Routledge Courzon.

Bulut, Christiane (2006). "Syntactic traces of Turkic-Iranian continuity," in Johanson, Lars and Christiane Bulut, editors. [Turkic-Iranian contact areas: historical and linguistic aspects](#).

Wiesbaden, Germany: Otto Harrassowitz Verlag.

Caferoğlu, A. and Doerfer, G. (1959). "Das aserbajdschanische," in *Philologiae Turicae Fundamenta* 1, 280-307. Wiesbaden.

Campbell, George L. and King, Gareth (2013). *Compendium of the world's languages*. New York: Routledge.

Cash, Jennifer, John Eidson, Bettina Mann, and Martin Ramstedt (2012). [Max Planck Institute for Social Anthropology report 2010 – 2011, volume I](#) p. 21. Halle/Saale: Max Planck Institute for Social Anthropology.

Csernyei, Tamás. Professor of Turkic Linguistics. Hacettepe University, Ankara, Turkey. Personal communication. January 2014.

Demir, Sema Alsan. Linguistics professor at Bashkent University, Ankara, Turkey. Personal communication. December 2013, January 2014.

Dirks, William. (2005). *O'zbekcha / Inglizcha Lug'at (Uzbek / English Dictionary)*, p. 64. Tashkent: The Central Asian Heritage Group.

Dmitrieva, L. V. (1997). "Barabinskikh tatr iazyk," in *Iazyki mira - Tyurkskie iazyki*, p. 199-205. Moscow: Indrik.

Doerfer, Gerhard (1969). "Die Turksprachen Irans." *Türk Dili Araştırmaları Yıllığı Belleten* 13-23.

Doerfer, Gerhard (1971). *Khalaj materials*. Bloomington, Indiana: Indiana University Publications.

Doerfer, G. (December 15, 1988). "[Bayat](#)." *Encyclopaedia Iranica*.

Doerfer, Gerhard (1989). "Die kopula im Chaladsch," in Heissig, W. and K. Sagaster, editors. *Gedanke und wirkung*. Festschrift

zum 90. Geburtstag von Nikolaus Poppe, pp. 106-12. Wiesbaden, Germany: Otto Harrassowitz Verlag.

Doerfer, Gerhard and Hesse, Wolfram. (1993). Chorasantürkisch. Wörterlisten-Kurzgrammatiken-Indices, 6-7. Wiesbaden.

Doerfer, Gerhard and Hesse, Wolfram. (1989). "Afšār-e Nānakčī," in Südoghische Materialien aus Afghanistan und Iran, pp. 71-123. Wiesbaden.

Doerfer, Gerhard (1998). "[Turkic languages of Iran](#)," in Johanson, Lars and Éva Á Csató, editors. The Turkic languages. London: Routledge.

Doerfer, G. (August 18, 2011). "[Azerbaijan VII. Azeri Turkish](#)." *Encyclopaedia Iranica*.

Dokuzlar, Uygur. Crimean Tatar speaker. Personal communication. April 2010.

Dolatkhah, Sohrab (2012). "Elements for a grammar of Kashkay, a Turkic language of Iran (unpublished PhD dissertation)." Paris: Ecole Pratique des Hautes Etudes.

Dolatkhah, Sohrab. Post-doctorate researcher at Ecole Pratique des Hautes Etudes at the Sorbonne, Paris. Personal communication. December 2013.

Donahoe, Brian Robert (2003). [A line in the Sayans: History and divergent perceptions of property among the Tozhu and Tofa of South Siberia](#). PhD dissertation. Bloomington, Indiana: Department of Anthropology, Indiana University.

Dougherty, William H. (Undated). "[The Tuvinian Autonomous Soviet Socialist Republic](#)."

Dwyer, Arienne M. (May 21, 2001). "[Salar grammatical sketch](#)." *DOBES: Documentation of Endangered Languages*. Hanover, Germany: Volkswagen Foundation.

Dwyer, Arienne M. Professor, Department of Linguistic Anthropology, University of Kansas, Lawrence, Kansas. Personal communication. August 2015

Dybo, Anna (2006). Khronologija Tyurkskikh jazykov i lingvisticheskiye kontakty rannikh Tyurkov ([The chronology of the Turkic languages and the linguistic contacts of the early Turks](#)). Moscow.

Eker, Süer and Kasapoglu, Çengel Hülya (September 20-24, 2009). "[Kipchak Turkic as a part of the Balkans and Eastern Europe history-geography](#)." Lucrările Simpozionului Internațional Cartea. Bucharest: Editura Biblioteca Bucureștilor.

Eker, Süer. Professor, Department of Turkish Language and Literature, Bashkent University, Ankara, Turkey. Personal communication. November-December 2013.

Eker, Süer. Professor, Department of Turkish Language and Literature, Bashkent University, Ankara, Turkey. Personal communication. August 2014.

Eker, Süer. Professor, Department of Turkish Language and Literature, Bashkent University, Ankara, Turkey. Personal communication. April and September 2015.

Enduring Voices (December 13, 2010). "[Monchak Tuvan goat color and pattern naming system](#)." Youtube.

Eriksonas, Tomas (May 23-26, 2012). "[On the phonology and morphology of the Taiga-Sayan Tuha language](#)." Presented at the 4th International Turkish Research Symposium. Hacettepe University, Ankara.

Eriksonas, Tomas. Graduate student at the Institute for Asian and African Studies, University of Helsinki, Finland. Personal communication. December 2013.

Erdal, Marcel and Nevskaya, Irina (2006). "[Introduction](#)," in

Exploring the eastern frontiers of Turkic. Wiesbaden, Germany: Otto Harrassowitz Verlag.

Eshoo, Jacklin. South Azeri speaker. Personal communication. August 13, 2014.

Esla, Sam Poova. South Azeri and Turkish speaker. Personal Communication. July 17, 2014.

Ektazarov, Juliboy (Summer/Winter 2014). *Об огузских говорах с угрозой исчезновения в южном казахстане* (An endangered Oghuz dialect in southern Kazakhstan). Endangered Turkic Languages: Turkic Languages Volume III, Issue 4-5, in print.

Fitzhugh, William W., editor (2002). [2001-2002 Mongolia field report: Mongolia's Arctic connections: the Hovsgol Deerstone project](#), p. 37. Washington DC: Arctic Studies Center, Smithsonian Institution.

Fraenkel, Gerd. (1962). "Mutual intelligibility between Turkish of Turkey and Azerbaijani," in Nicholas Poppe, editor. American studies in Altaic linguistics, Bloomington: Indiana University.

Funk, D. A. (2005a) [Worlds of shamans and storytellers \(A study of Teleut and Shors materials\): Résumé](#). Moscow: Nauka.

Funk, D. A. (2005b). "[The language of the Chulym Turks](#)," in Funk, D. A., V. Y. Gusev and A. N. Bitkeeva, editors. Endangered languages of indigenous peoples of Siberia. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Galiullina, Julia (September 26, 2011). [Red book of dying languages](#). Moscow: The Voice of Russia.

Garkavets, Alexander (Undated). "[North Azovian Urums - Greeks speaking Turkic language](#)." *Christus Rex*.

Georg, Stefan (2003). "[The gradual disappearance of a language family: The case of Yeniseian](#)," in Janse, Mark and Sijmen Tol, editors. Language death and language maintenance: Theoretical, practical and descriptive approaches, p. 96. Amsterdam: John Benjamins Publishing.

Grimes, Barbara F., editor (1996). "[Languages of Iran](#)," in Ethnologue: languages of the world, thirteenth ed. Dallas: Summer Institute of Linguistics.

Grimes, Barbara F., editor (2000). "[Ili Turki](#)," in Ethnologue: languages of the world, fourteenth edition. Dallas, Texas: SIL International.

Grimes, Barbara F., editor (2000a). "[Azeri](#)" in Ethnologue: languages of the world, fourteenth edition. Dallas, Texas: SIL International.

Güzel, Sinan. Professor, Department of Turkic Languages and Literature, Izmir Katib Chelbi University, Izmir, Turkey. Personal communication. April 2015.

Hahn, Reinhard F. (1998). "[Uyghur](#)," in Johanson, Lars and Éva Á. Csató, editors. The Turkic languages. London: Routledge.

Harrison, K. David (Undated a). "[Fieldwork: Monchak 2002](#)." K. David Harrison, PhD. webpage at Swarthmore University.

Harrison, K. David (Undated b). "[Fieldwork: Tsengel Tuvan 2002](#)." K. David Harrison, PhD. webpage at Swarthmore University.

Harrison, K. David (2002). "[Tofa](#)," in Encyclopedia of the world's minorities. Chicago: Fitzroy Dearborn.

Harrison, K. David (2006). "[South Siberian sound symbolism](#)," in Languages and prehistory of Central Siberia. Edited by Vajda, Edward J. Amsterdam: John Benjamins Publishing.

Harrison, K. David (2007). [When languages die: the extinction of the world's languages and the erosion of human knowledge](#). Oxford: Oxford University Press.

Howarth, Henry H. (February 10, 1883). "The Turkish Elements in Magyar," in *The Academy: A Weekly Review of Literature, Science and Art*, January-June 1883, Volumes 23-24. London: Alexander and Shephard.

Hu, Zhen-hua and Imart, Guy (1987). *Fu-Yü Gïrgïs: A tentative description of the easternmost Turkic language*. Bloomington, Indiana: Indiana University Research Institute for Inner Asian Studies

Humphreys, Andrew and Mits, Krista (1991a). "[Altaics](#)," in *The red book of the peoples of the Russian Empire*. Tallinn: NGO Red Book.

Humphreys, Andrew and Mits, Krista (1991b). "[Crimean Tatars](#)," in *The red book of the peoples of the Russian Empire*. Tallinn: NGO Red Book.

Humphreys, Andrew and Mits, Krista (1991c). "[Dolgars](#)," in *The red book of the peoples of the Russian Empire*. Tallinn: NGO Red Book.

Humphreys, Andrew and Mits, Krista (1991d). "[Lithuanian Tatars](#)," in *The red book of the peoples of the Russian Empire*. Tallinn: NGO Red Book.

Imin, Tursanjan. Professor, Institute of Uyghur Culture and Literature, Northwest University for Nationalities, Lanzhou, Gansu, China. Personal communication. December 2013.

Jankowski, Henryk (2000). "Crimean Tatars and Noghais in Turkey". *Türk Dilleri Araştırmaları [Studies on the Turkic Languages]* 10: 113-131. Ankara: Sanat Kitabevi

Janhunen, Juha and Salminen, Tapani (December 18-28, 1993). [UNESCO red book on endangered languages: Northeast Asia](#).

Johanson, Lars (August 18, 2011). "[Azerbaijan IX. Iranian Elements in Azeri Turkish](#)." *Encyclopaedia Iranica*.

Johanson, Lars (2001). "[Discoveries on the Turkic linguistic map](#)." Stockholm: Swedish Research Institute in Istanbul Publication 5.

Johanson, Lars (2005). "[Introduction: Converging codes in Iranian, Semitic And Turkic](#)," in Csató, Éva Ágnes, Bo Isaksson, and Carina Jahani, editors. Linguistic convergence and areal diffusion: Case studies from Iranian, Semitic and Turkic 14. New York: Routledge Courzon.

Johanson, Lars (2006). "[On the roles of Turkic in the Caucasus area](#)," in McMahon, April, Yaron Matras and Nigel Vincent, editors. Linguistic areas, pp. 160-181. London: Palgrave Macmillan.

Johanson, Lars and Ragagnin, Elizebetta (2006). "Central Asia and Mongolia," in [Sociolinguistics: An international handbook of the science of language and society](#), Volume 3. Ammon, Ulrich, Norbert Dittmar, Klaus J. Mattheier, and Peter Trudgill, editors. Berlin: Walter de Gruyter.

Kaluzynski, Stanislaw (1962). Mongolische elemente in der jakutischen sprache. Warsaw, 'S-Gravenhage: Panstwowe Wydawnictwo Naukowe, Mouton & Co.

Karakoc, Birsal (June 2014). "[Az konuşurlu bîr Türk dîlî: Nogayca ve Nogaylar üzerîne](#). [A Turkic language with few speakers: Nogay language and Nogay People]." *Nogai Ногайцы*.

Khalid, Adeeb. 1998. The Politics of Muslim Cultural Reform: Jadidism in Central Asia, p. 190. Berkeley: University of California Press.

Kharlampovich, Konstantin Vasil'evich and Haney, James Lawton (translator and editor) (2001). Archimandrite Makarii Glukharev - founder of the Altai Mission. Lewiston, New York: Edwin Mellon

Press.

Knüppel, Michael (April 15, 2010). "[Turkic languages of Persia: An overview](#)." *Encyclopaedia Iranica*.

Kosterin, O. E. and Zaika, V. V. (2003). "[Odonatological expeditions to the Tyva Republic \(Tuva\) 2000-2002](#)." *IDF-Report* 5, 1-44.

Lement, Victoria C. (2013). To-Z - "[To-Z: Culture of Türkmenistan. Türkmenistan](#)." *Countries and Their Cultures*.

Lenskaya, V. M. (2013). "[О Словообразовании В Чулымско-Тюркском Языке \(On Chulym Turkic word formation\)](#)." *Tomsk State Pedagogical University Bulletin* 10, 128-131.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009a). "[Balkan Gagauz Turkish](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009b). "[Bashkir](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009c). "[Chuvash](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009d). "[Dolgan](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009e). "[Ili Turki](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009f). "[Karachay-Balkar](#)," in *Ethnologue: Languages of the world*, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009g). "[Karaim](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009h). "[Khakas](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009i). "[Khalaj](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009j). "[Kirghiz](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009k). "[Kumyk](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009l). "[Northern Altai](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009m). "[Northern Uzbek](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009n). "[Salar](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009o). "[Salchug](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009p). "[Southern Altai](#)," in Ethnologue: Languages of the

world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009q). "[Southern Uzbek](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2009r). "[Tatar](#)," in Ethnologue: Languages of the world, sixteenth edition. Dallas, Texas: SIL International.

Lewis, M. Paul, Gary F. Simons and Charles D. Fennig, editors (2014). "South Azerbaijani" in Ethnologue: Languages of the world, seventeenth edition. Dallas, Texas: SIL International.

Living Tongues Institute for Endangered Languages (2007-present a). "[Altai-Sayan language and ethnography project](#)." Salem, Oregon: Living Tongues.org.

Living Tongues Institute for Endangered Languages (2007-present b). "[The Ös documentation project](#)." Salem, Oregon: Living Tongues.org.

Mair, Victor H. Professor of Chinese at the University of Pennsylvania. Personal communication. January 2009.

Maneshev, Ackhat. Sibir Tatar speaker, Russian Federation. Personal communication. December 2013.

Matthews, W. K. 2013 (1951). Languages of the USSR. Cambridge, UK: Cambridge University Press.

Menges, Karl H. (1947). Qaraqalpaq grammar. Translated from German by Leora P. Cunningham. New York: King's Crown Press.

Menges, K. H. (1951). "Research in the Turkic dialects of Iran (Preliminary report on a trip to Persia)." *Oriens* 4, 273-79.

Minorities At Risk Project (December 31, 2006). "[Assessment for Kumyks in Russia](#)." College Park, MD: University of Maryland

Center for International Development and Conflict Management.

Minorsky, V. (1940). "The Turkish dialect of Khalaj." *Bulletin of the School of Oriental Studies*.

Mongush, D. A., and Sat, Sh. Ch. (1967). "[Development of the Tuvan language during the Soviet period](#)," 42-46. Xyzyl: Tyvanyň Nom Ündürer Cheri.

Mongush, D. A. (2012a). "[The Tophas and the Soyots: Historical and ethnographic essay](#)." *The New Research of Tuva* 2.

Mongush, D. A. (2012b). "Tuvans Xinjiang: Twenty years later." *The New Research of Tuva* 4.

Moseley, Christopher (ed.) (2010a). "[Alabugat Tatar](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010b). "[Baraba Tatar](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010c). "[Bashkir](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010d). "[Dolgan](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010e). "[Kamas](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010f). "[Kamas Turk](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010g). "[Karachay Balkar](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010h). "Karagash," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010i). "[Khakas](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010j). "[Northern Altai](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Moseley, Christopher (ed.) (2010k). "[Yakut](#)," in Atlas of the world's languages in danger, 3rd ed. Paris: UNESCO Publishing.

Mote, Victor L. (2004). "[Altai](#)." Encyclopedia of Russian history. *Encyclopedia.com*.

Mudrak, O. A. (2002). Ob utochnenii klassifikatsii tyurkskikh jazykov s pomosch'ju morphologicheskoy lingvostatistiki [On the clarification of the Turkic languages classification by means of morphological linguostatistics]. Sravnitel'no-istoricheskaja grammatika tyurkskikh jazykov. Regionalnyye rekonstruktsii. Moscow.

Németh, Michał (2011). "A different look at the Lutsik Karaim sound system (from the second half of the 19th century)," in *Studia Linguistica Universitatis Jagellonicae Cracoviensis*, Elżbieta Mańczak-Wohlfeld, ed. Vol. 128. Warsaw: Jagellonian University Press.

Nevskaya, I. A. (2005). "[Teleut](#)," in D. A. Funk, V. Y. Gusev and A. N. Bitkeeva, editors. Endangered languages of indigenous peoples of Siberia. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Nixon, Melody (September 23, 2005). "[About today's Teleuts. Teleuts - the hidden people in Siberia](#)." *Sibirga SNTC Native News and Culture*. Ontario, Canada: Miighan-Kurt Company.

Oberling, Pierre (December 15, 2008). "Khorasan i. Ethnic groups." *Encyclopaedia Iranica*.

Ögel, Bahaeddin (1989). Türk mitolojisi - Kaynakları ve açıklamaları ile destanlar [Turkic mythology]. Ankara: Türk Tarih Kurumu.

Opoudjis, Aka (April 10, 2004). "Ἀλγηνιστεύκοντος [[The status of Urum](#)]." Melbourne, Australia.

Pakendorf, Brigitte (2007). Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives. PhD dissertation, Universiteit Leiden. Utrecht: LOT.

Podol'skii, Baruch (1985). [A Greek Tatar-English glossary](#). Wiesbaden, Germany: Otto Harrassowitz Verlag.

Potapov, L. P. (1969). [Ethnic composition and origin of Altaians - historical ethnographical essay](#). Leningrad: "Science" Publishing House, Leningrad Branch.

Ragagnin, Elisabetta (2006). "[The position of Dukhan among the Tuvan dialects](#)," in Erdal, Marcel and Irina Nevskaya, editors. Exploring the eastern frontiers of Turkic. Wiesbaden, Germany: Otto Harrassowitz Verlag.

Ragagnin, Elisabetta (2009). "[A rediscovered Lowland Tofan variety in Northern Mongolia](#)." *Turkic Languages* 13, 225-245.

Ragagnin, Elisabetta (2012). "[Taiga vs. Steppe Sayan Turkic varieties in Northern Mongolia](#)," in Dilleri ve kültürleri yok olma tehlikesine maruz türk toplulukları konulu 4. Uluslararası Türkiyat Araştırmaları Bildiriler [The 4th International Turkish Research Symposium: Turkic communities whose languages and cultures are endangered], Hazırlayanlar, Yayına, Mikail Cengiz and Tefik Orçun Özgün, editors. Ankara: Hacettepe University Hospitals Printing House.

Rassadin V. I. (1978). Morfologija tofalarskogo jayzka v

sравnitel'nom osveščanii [The morphology of Tofa in comparative perspective]. Moscow: Izdatel'stvo 'Nauka'.

Rassadin, V. I. (2005a). "[The Soyot language](#)," in Funk, D. A., V. Y. Gusev and A. N. Bitkeeva, editors. Endangered languages of indigenous peoples of Siberia. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Rassadin, V. I. (2005b). "[The Tofalar Language](#)," in Funk, D. A., V. Y. Gusev and A. N. Bitkeeva, editors. Endangered languages of indigenous peoples of Siberia. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Rassadin, Valentin Ivanovich (2013). "[The problems of revival and preservation of languages of some small Turkic peoples of South Siberia \(On an example of Tofa Soyot languages\)](#)." *Middle-East Journal of Scientific Research* 17:4, 546-550.

Radloff, Wilhelm (1908). Die jakutische Sprache in ihrem Verhältnisse zu den Türksprachen. Mémoires de l'Académie Impériale des Sciences de St-Pétersbourg. VIII Série. St. Petersburg: Classe Historico-Philologique.

Róna-Tas, András, Árpád Berta and László Károly (2011). *Tucologica #84: West Old Turkic: Turkic loanwords in Hungarian*. Wiesbaden, Germany: Harrassowitz Verlag.

Roos, Marti; Hans Nugteren, Zhong Jìnwén (July 28, 2009.) [The Turkic language family](#).

Ryzhikova, Tatiana, Iraida Selyutin, Nikolai Urtegeshev, and Andrey Letyagin (2012). "[MRI verilerine göre Sibirya Tatarların gırtlak sesletimleri арингализация в языке сибирских татар \(по данным мрт\) \[Pharyngealization in the Siberian Tartars' language \(MRI investigation\)\]](#)." *Karadeniz: An International Quarterly Journal of Social Science* 5:19, 503-51.

Salminen, Tapani (2007a). "[Europe and North Asia](#)," in Encyclopedia of the world's endangered languages, 211–280, edited by Christopher Moseley. London & New York: Routledge.

Salminen, Tapani (2007b). "[Endangered languages of Europe](#)," in Language diversity endangered, edited by Matthias Brenzinger. Berlin: Mouton de Gruyter.

Salminen, Tapani. Professor of Finno-Ugrian Language Studies, Department of Finno-Ugrian Languages and Cultures, University of Helsinki, Helsinki, Finland. Personal communication. September 2014.

Sengupta, Anita (2003). The formation of the Uzbek nation-state: A study in transition. Lanham, Maryland: Lexington Books.

Sierakowska-Dyndo, Jolanta (2013). [The boundaries of Afghans' political imagination: the normative-axiological aspects of Afghan tradition](#). Newcastle upon Tyne: Cambridge Scholars Publishing.

Skopeteas, Stavros, Violeta Moisidi, Eleni Sella-Mazi and Efy Yordanoglu (2011). "[Urum documentation project](#)." Bielefeld, Germany: University of Bielefeld.

Skopeteas, Stavros. Professor of Linguistics at University of Bielefeld, Bielefeld, Germany. Personal communication. December 2013.

Stachowski, Marek (2011). "[Siberian languages in contact, 1: Collective numerals in Yakut, Dolgan, Tuvian, Tofalar and some other Turkic languages](#)." In *Studia Linguistica Universitatis Jagellonicae Cracoviensis*, Vol. 128. Elżbieta Mańczak-Wohlfeld, ed. Warsaw: Jagellonian University Press.

Ståhlberg, Sabira (July 6, 2013). "[Tatars in Finland](#)." *Sabirien*.

Stein, Marc Aurel (1912). Ruins of Desert Cathay: Personal Narrative of Explorations in Central Asia and Westernmost China, Volume 1, , p. 342. London: Macmillan.

Stukova, Natalia (2006). "[The sociolinguistic situation in Mountain Shoria. Exploring the eastern frontiers of Turkic](#)," in Erdal, Marcel and Irina Nevskaya, editors. Exploring the eastern frontiers of Turkic. Wiesbaden, Germany: Otto Harrassowitz Verlag.

Tatár, Maria Magdolna (1979). "[The Khotons of Western Mongolia](#)." *Acta Orientalia Hungarica XXXIII* 1, 1-39.

Tazranova, A. R. (2005). "[The Chelkan language](#)," in Funk, D. A., V. Y. Gusev and A. N. Bitkeeva, editors. Endangered languages of indigenous peoples of Siberia. Moscow: Department of Northern and Siberian Peoples, Institute of Ethnology and Anthropology, Russian Academy of Sciences.

Tehrani, Jamshid J., Mark Collard, and Stephen J. Shennan. (December 12, 2010). The cophylogeny of populations and cultures: Reconstructing the evolution of Iranian tribal craft traditions using trees and jungles. *Phil. Trans. R. Soc. B*.

Tekin, Talat (1978). "[Türk dilleri ailesi \(Turkic family of languages\)](#)." *Türk Dili* 37:318, 173-183.

Tekin, Talat (1990). "[A new classification of the Turkic languages](#)." *Tr'iric Dilleri Arastirmlari* 5, 18.

Tenishev, E. and Todayeva, B. (1966). Yazyk zhyoltykh uyghurov [The language of the Yellow Uyghurs]. Moscow.

Tishkov, V. A., V. V. Stepanov, D. A. Funk, and O. I. Artemenko (2009). [Status of and support for linguistic diversity in the Russian Federation](#). Moscow: Institute of Ethnology and Anthropology of the Russian Academy of Sciences, Ministry of Education and Science, Russian Federation, Centre for National Issues related to Education, Federal Institute for the Development of Education.

Todoriki, Masahiko (July-September 2010). "[Tuwa-ren: The emerging ethnic identity of the Altai-Tuvans in Xinjiang](#)." *Journal*

Of Eurasian Studies 2:3, 91-103.

UCLA Language Materials Project a (Undated). "[Tatar](#)." Los Angeles: UCLA International Institute Center for World Languages.

UCLA Language Materials Project b (Undated). "[Yakut](#)." Los Angeles: UCLA International Institute Center for World Languages.

Vovina, Olessia (September 2006). "[Islam and the creation of sacred space: The Mishar Tatars in Chuvashia](#)." *Religion, State & Society* 34:3, p. 258.

Wind of Mongolia (November 21, 2011). "[Trekking in Altai - Mongolia - Altai Tavan bogd - trekking tour \(11 days, 11 nights\)](#)." *Extreme Mongolia*.

Windfuhr, Gernot (March 29, 2012). "[Iran VII. Non-Iranian languages \(7\) Turkic languages](#)." *Encyclopaedia Iranica*.