

Turkic: A family portrait

Lars Johanson (Mainz)

A family and its members

In this paper, I will invite Turkic, the family to which Turkish belongs, to sit down for a quick portrait, a snapshot.

The Turkic-speaking world is comprehensive. It extends from the Southwest, Turkey and her neighbors, to the Southeast, to Eastern Turkistan and farther into China. From here it stretches to the Northeast, via South and North Siberia up to the Arctic Ocean, and finally to the Northwest, across West Siberia and East Europe.

The area comprises a great number of different languages. The regions in which Turkic languages are spoken include Anatolia, Azerbaijan, the Caucasus region, Iran, Iraq, Afghanistan, the immense areas of West and East Turkistan, South, North and West Siberia and the Volga region. In the past, the Turkic-speaking world also included compact areas in the Ponto-Caspian steppes, the Crimea, the Balkans, etc.

There are currently about twenty Turkic standard languages, the most important ones being Turkish, Azeri, Turkmen, Kazak, Karakalpak, Kirghiz, Uzbek, Uyghur, Tuvan, Yakut, Tatar, Bashkir and Chuvash.

Due to their development at the end of the twentieth century, after the breakdown of the Soviet Union, many Turkic languages have acquired increased political importance as national languages of a set of new autonomous states.

A language family

According to the criteria provided by traditional comparative linguistics, Turkic is a language family, a group of related languages descended from a single ancestral language. We can clearly prove that Turkish, Uzbek, Yakut, Chuvash and others are interrelated, just as we can prove that Swedish, Lithuanian, Greek and Armenian are relatives within the Indo-European family. There was a period in which scholars had doubts about the affiliation of Chuvash, but today it is clear that this is also a genuine Turkic language.

There are thus reasons to assume a Turkic protolanguage and its subsequent differentiation, which may be represented in the form of a genealogical tree, a family tree, a pedigree, in common metaphorical parlance. But a linguistic family is somewhat different from a family of human beings of common ancestry. The latter go back to two parents, but a linguistic family is a single-parent family. It only includes mothers and daughters. It comprises codes with one single parent each, that is 'mother codes'. It is fashionable today to speak of so-called mixed languages, which are claimed to have two parents, but these are, if they exist, utterly marginal phenomena.

Family members

Let us range the modern Turkic languages together to sit for a family portrait. What do we get? About 20 languages in the political sense, established by political decisions, connected with a certain history, culture, literature, political structure, etc., and endowed with their own institutions. Most of them were created in the 20th century through relatively arbitrary acts of language policy. Many of them emerged as rather loose bundles of more or less interrelated dialects. Each has some kind of standard language, and related non-standard varieties overroofed by the standard. There are also unroofed varieties outside the area of validity of the standard languages.

Classifications according to political and purely linguistic criteria may yield different results. Thus, East Anatolian Turkish is closer to Azeri with respect to genealogical proximity and similarity than to the rather strongly Istanbul-based national language of Turkey. Karakalpak is linguistically a dialect of Kazakh, but a language in the political sense. Dolgan is a dialect of Yakut according to the interintelligibility criterion, but its speakers consider it a language in its own right.

Branches and subdivisions

Most members of the family can be classified as varieties of higher groups, primary branches: the Oghuz (SW), the Kipchak (NW) branch, the Uyghur-Karluk (SE) branch. They have their specific features and it is mostly easy to decide which branch a given modern Turkic language belongs to. Thus Turkish is a variety of

Oghuz, which is a variety of Turkic. Khalaj, spoken in Central Iran, and Chuvash, spoken in the Volga area, constitute special branches.

Members of language families, unlike individuals in a human family, divide into varieties and subvarieties. Through permanent differentiation, more specialized kinds of spoken Turkic have emerged. The family tree branches out into geographic varieties: dialect groups, regional dialects and basic local dialects. And in social varieties, more or less important in terms of prestige. Applying a two-dimensional dialectology, we may delimit *lects*, geographically and socially distinctive varieties that combine distinctive features belonging to particular areas and layers. They contrast horizontally with the neighbours, and vertically with other social varieties, including standard varieties. Each has a certain range of validity, communicative range. It is of course impossible to portray these myriads of varieties.

Sign language

Let me briefly mention a very specific problem of classification of Turkic. There is a largely unknown language in Turkey, the sign language of the deaf community, officially endorsed by new legislation in 2005. It is not a representation of spoken Turkish “on the hands”. It has its very own structure and is in many respects radically different from Turkish, in morphology and, to a lesser extent, syntax. Unlike spoken Turkish, it has complex classificatory constructions, numeral incorporation and verb agreement with both subject and object. It has no direct counterpart to the tense and case systems of spoken Turkish. The Turkish sign language is certainly not Turkic in the same sense as spoken Turkish is Turkic. But there are specific connections between the two languages which should be described. Incidentally, it is interesting to note that there was an elaborate Turkish sign language in use at the Ottoman court already in the 16th century.

Field work

Parts of the Turkic linguistic world have so far been insufficiently investigated. Field work necessary in many fields.

Let me briefly mention the Swedish tradition of field research in the Turkic world. It begins with the research carried out by Swedish officers of Charles XII,

who had fallen into captivity in Siberia after the battle of Poltava (1709). With his zealous scientific activity in Siberia, his discoveries of inscriptions and manuscripts, Filip Johan von Strahlenberg (1676-1747) stands out as a kind of progenitor of Turcology or even Uralo-Altaistics. His monumental work "Das Nord- und Ostliche Theil von Europa und Asia" was printed in Stockholm in 1730.

On the upper course of the Yenisey, Strahlenberg and others had found burial-places and stone inscriptions written in an enigmatic script whose letters were similar to Nordic runes. Later, the language of the inscriptions turned out to be what has been called "Old Turkic". The Turkic so-called "Runic script" was thus known as early as at the beginning of the eighteenth century, though it was to remain undeciphered until the end of the nineteenth century. The greatest discovery in the history of Turcology was made in the Orkhon valley in today's Mongolia in the summer of 1889. A Russian scientific expedition discovered a number of big stone stelæ covered with inscriptions. The texts were written with signs of the same runiform type that was already known from stones found by Strahlenberg and others. The discovery was reported very quickly, and the learned world began to take intense interest in the problem of the "runes". On December 15, 1893, the well-known Danish professor of comparative linguistics Vilhelm Thomsen announced that he had succeeded in deciphering the enigmatic script. It was suddenly possible to read East Old Turkic texts of the eighth century dedicated to the rulers of the Turk empire and glorifying their military achievements.

In the twentieth century, Swedish Turcology came to play a leading part in the investigation of the varieties spoken in Eastern Turkistan. The research was started by Gustaf Raquette, who had spent many years as a medical missionary in Yarkand and Kashgar and, after his return to Sweden, took up a lectureship at the University of Lund. The research was continued by his pupil Gunnar Jarring, who wrote the dissertation "Studien zu einer osttürkischen Lautlehre" which in 1933 brought him a position as a university lecturer at the university of Lund. It was followed by a set of publications, the outcome of strenuous field work, e.g. texts from regions in Chinese and Afghan Turkistan that soon afterwards became inaccessible. Gunnar Jarring is internationally recognized as a pioneering explorer of unknown Turkic dialects in Central Asia.

In the last decades, much important documentary work has been carried out. Gerhard Doerfer's research in Iran, which has opened up new ground, is a splendid example of this. Let me also mention Éva Csató's investigations on the Karaim varieties spoken in Lithuania and Halich and Elisabetta Ragagnin's current fieldwork on Dukhan, a hitherto undocumented variety of Sayan Turkic, spoken in Northern Mongolia.

Today we are facing new exciting possibilities for field work. The need for linguistic documentation is great. We not only need data from well-established Turkic languages, but also from less known vernacular varieties, peripheral languages, languages strongly influenced by contact, isolated languages displaying both archaic and innovative features, etc. Field research may bring about further important data, which may again lead to considerable re-evaluations in Turkic linguistics. Linguistic documentation is an urgent task that is best carried out in international cooperation. It is exciting, particularly for young people, to take active part in documenting languages and collecting new primary data. We still need linguistically trained scholars who can produce empirically adequate and theoretically meaningful research.

Endangerment

Some of today's Turkic languages are endangered, or at least potentially endangered. The endangerment starts when young generations begin to pay less attention to their primary code and switch over to a dominant code because they find it more attractive and prestigious. Currently, increasing endangerment and death of languages is observed all over the world, a development that leads to mass death of languages and will, like other kinds of globalization, extinguish variation in an irrevocable way. It is important to try to document endangered Turkic languages while it is still possible to do so. The Hans Rausing Endangered Languages Project at SOAS in London is dedicated to documentary linguistics, trying to play a role in attacking the problem of language endangerment, by providing training, archiving, publishing, and funding of projects. The services of this project may become useful to some of the Turkic communities whose languages are endangered and to the researchers who work with them.

Copying

The members of the family we are trying to portray have emerged as a result of internal development and complex contact processes with copying of foreign elements. Speakers of Turkic have taken over copies into their own primary code and non-Turkic speakers shifting to Turkic have carried over copies from the own primary code into their Turkic secondary code. Because of the unique mobility of Turkic-speaking groups contact-driven developments have been especially important.

The influence of Turkic on other languages has been equally great. One example is the Turkish influence on the Romani dialects of the Balkans, especially the extensive and varied copying and integration of grammatical structures. André Hesselbäck, Uppsala, has dealt with Turkic code-copying influence on the Finno-Ugric language Mari. Don Stilo is working on shared phonological, grammatical and lexical features in languages of the South Caucasus, eastern Turkey, Northern Iran and Northern Iraq. What he refers to as the ‘Araxes Linguistic Area’ involves heavy contact phenomena among five different language families: Turkic, Kartvelian, Indo-European, Semitic and Northeast Caucasian.

Because of contact-induced change, genealogically unrelated varieties have become more similar to each other, e.g. Uzbek Turkic and Tajik Persian. But we do not find mixed languages. Classical written Ottoman texts may contain 90 % loanwords, but they still have to be defined as Turkic and not as Arabic or Persian. The same is true of languages such as Karaim and Gagauz, strongly influenced by Slavic. We find excessive copying of phonology, syntax, lexicon, etc., but never a turnover to the languages copied from.

Heavy code-copying does not cause code replacement. Shift does not mean successive transition from one code to another through intermediary stages characterised by increasing copying. Speakers of a dominated code do not take over larger and larger parts of a dominant code, until they end up speaking the dominant code instead of the dominated one. Increasing influence on a dominated code does not lead to its abandonment. Languages are not abandoned for structural reasons. The decisive factors for shift are social in nature. Languages are abandoned because

they are no longer handed down. If social pressure leads to a negative attitude towards a recessive code, its speakers may avoid transmitting it to their children, who acquire it incompletely and, at best, grow up as semi-speakers.

In the development of Turkic we always have to reckon with contact situations in which mutually intelligible varieties have met and influenced each other. This was the normal situation in tribal confederations with their mobile heterogeneous groups. The encounters led to changes and the emergence of modified varieties. There were dominant varieties, for example koinés, which may have copied extensively from other varieties but maintained their position and survived as such. Contacts between mutually intelligible codes can lead to glottogenesis; a conglomerate of varieties of different origin may develop into a more homogeneous variety. Today's Kashgari, spoken in Iran, may probably serve as a model for similar situations in the history of Turkic.

One Turkic language?

Within the Turkic-speaking world, there are still ideologically motivated, linguistically nonsensical discussions on which idioms should be termed 'languages' or 'dialects'. Many Turkish scholars refer to the family as one language, *dil*, and to each member as a dialect, *lehçe*. As always in questions of terminology, the criteria must be defined. The linguistic term 'dialect' is used for varieties that stand in a relationship of mutual intelligibility to others. And this criterion does not apply to the idioms we are trying to portray. The family members cannot chat freely with each other. Neighbours may do some small talk. But no member can converse directly with the whole family. Maybe this is even an essential condition for a happy family life.

It is also a misunderstanding that there was once one uniform Turkic whose unity was destroyed in modern times, i.e. that the differences within the Turkic family arose as a result of modern language policy. This idea is caused by confusion of spoken and written languages. It is true that, for example, the validity of Chaghatay as a superregional literary language was reduced and finally destroyed by national languages such as Uzbek, Uyghur, Tatar, Turkmen, etc. This is

something else and does not contradict the fact that the differentiation of spoken Turkic began very early.

Creation of a family language

The hope to create a language for the whole family seems to be futile. At the beginning of the 20th century, Tatar and Turkestanian intellectuals began to discuss whether a common written language should replace the different regional languages. Advocates of a cultural panturkism wanted to create a common standard as a unifying bond. Some of them imagined a common means of communication for all Turks “from the boatmen of Istanbul to the shepherds of Eastern Turkestan”. There was, however, disagreement to what extent this language should lean on Tatar or on Ottoman. The Tatars had developed a written standard close to spoken Kazan Tatar, which also exerted considerable influence in Turkestan. One promoter of a common language was Isma’il Bey Gaspīralī, who published the newspaper “Terjiman” in the Crimea. His paper addressed all Turkic-speaking groups of Russia, thus aiming for as wide a communication radius as possible. In order to secure a broad comprehensibility it used numerous Ottoman elements. Consequently, its language was criticized for being “artificial”, a kind of Turkic Esperanto. The unification efforts remained fruitless for political reasons. It also became impossible to maintain the old supraregional languages and elaborate them as modern standards.

Today there are new dreams of a common Turkic, ideas of removing differences between the Turkic languages in order to create a unified language that might get official status on international platforms, like, for instance, Chinese or Arabic. These are hardly realistic ideas. The linguistic differentiation is an accomplished fact and irreversible. There is no “average” or “normal” Turkic to replace the “national” languages, and it has never existed.

Of course it would be possible to harmonize the standard languages, i.e. to reduce differences through coordinated language planning, creating unified systems of terminologies, etc. But even this meets insurmountable political obstacles. Particularism is still prevailing. The language policies of the new Turkic republics are guided by the wish to create state languages to help establish new national

identities for their populations. Not even the relatively realistic goals of harmonization the writing systems have been reached. Even the chance to harmonize terminologies seems to slip through the fingers of the reformers.

A wider family?

We all know from our personal lives that genealogical research can reveal wider family relations than ever imagined before. The question is: Does Turkic have relatives, i.e. is it part of a larger family? This issue has been discussed very emotionally and polemically for a long time. It has not been possible to prove that Turkic is related to Mongolic, Tungusic, etc. in the same sense that Germanic is related to Romance and Slavic, etc. On the other hand, nobody has been able to disprove the relatedness of the so-called Altaic languages.

Certain linguistic features are not easily replaced by take-over copying. The genealogical affiliation of a high-copying code may be determined by means of the elements that are generally least susceptible to being replaced. In a book on structural factors in Turkic contacts I have ventured to claim that the suffixes standing closest to the stem of a Turkic verb, namely those expressing actionality and diathesis, are most impervious to copying. Old morphemes expressing actionality and diathesis occurring next to the primary verbal stem display similarities in Turkic, Mongolic, Tungusic, Korean and Japanese. They might be remnants of a protolanguage that later split into different varieties subject to extensive copying of both kinds. Martine Robbeets, a Humboldt fellow at Mainz, has, in her research, focused on genealogical questions of ‘Altaic’. She has found striking phonetic and semantic similarities between morphemes occurring next to the primary verbal stem over a huge area from Turkish to Japanese. These similarities cannot simply be due to mere chance. Martine Robbeets has just published a critical book on the position of Japanese, asking the question: ‘Is Japanese related to Korean, Tungusic, Mongolic and Turkic?’. It seems to me that she is presenting good arguments for a possible ‘yes’.

Thus it is still not clear how large the family is that we should invite to sit for our family portrait.

