Werk

Titel: A new attempt to classify the Turkic languages (1)

Autor: Schönig, Claus

Ort: Wiesbaden

Jahr: 1997

PURL: https://resolver.sub.uni-goettingen.de/purl?666048797_0001|LOG_0018

Kontakt/Contact

Digizeitschriften e.V.
SUB Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen

info@digizeitschriften.de

A new attempt to classify the Turkic languages (1)

Claus Schönig

Schönig, Claus 1997. A new attempt to classify the Turkic languages (1). *Turkic Languages* 1, 117-133.

This attempt to classify the modern Turkic languages makes use of many data already known and sometimes used in earlier classification models. Older stages of Turkic, especially Old Turkic, are not neglected, but the author does not intend to integrate the historical steps of Turkic into this classification at this stage of its development. One important point is to include data that are not predominantly phonetic. Besides those, lexical, morphological and syntactic data are taken into consideration. Another important point is that a mere genetic classification is insufficient. Rather, we can observe developments of areal grouping during which Turkic languages of various genetic sub-branches form new territorial units and become involved in linguistic areal interaction. The article will be published in three parts.

1. Introduction

This is an attempt to classify the Turkic languages with respect to linguistic features, the relevance of which has been proven by comparative works about North East Turkic, Turkish and Kirghiz. It makes use of many data already known and used in earlier classification models, but sometimes puts them into new constellations and correlations.

My aim is to classify the modern Turkic units, let us say of the last 100 or 150 years. I will pay attention to historical stages of Turkic and try to connect—if possible—modern linguistic facts with older ones. But I do not intend to integrate the historical steps of Turkic into this classification at this stage of development of the proposed model. These steps should be investigated separately and synchronically before being incorporated.

An important point is not to include mostly phonetic data, as has been the case in former classification models; see, e.g., Arat (1929 and 1953),

Baskakov (1962), Benzing (1959b), Menges (1959a and 1968), Ramstedt (1957), Räsänen (1949 and 1953: 26–31), Samojlovič (1922), Doerfer (1985), Poppe (1965), Tekin (1995). All these models contain a number of important points, which I have tried to incorporate into my model. If such features belong to "common knowledge" in Turkology, I do not give bibliographical data concerning them.[1] Besides phonetic data, I have tried to make use of lexical, morphological and syntactic data. Phonetic features may be sufficient to individualize every single Turkic unit (even subdialects), but languages consist of more than sounds. If one wants to set up a classification which can be used to explain historical developments and interactions with other languages, one must find features in all the other fields of grammar.

Another important point is that a merely "genetic" classification is not sufficient. We have good reasons to assume that all the contemporary Turkic languages have not *directly* developed from a common ancestor. According to such a "tree model" this "proto-Turkic" ancestor would have split up into a certain number of sub-groups, which again would have split up into smaller groups and so on. The reason for this development would be that in separate parts of the ancestral unit some linguistic features change with different results while shifting from one time level to another. Such a transmission of features from one historical step to the next I call *genetic heritage*. But if we have to give up the idea that genetic heritage is the only way of diachronic transmission of linguistic features, we can no longer hang on to the model of a genetic "tree", leading back to one "proto-Turkic" unit, which split into more and more sub-units until finally reaching the status of today. This is only one component in the development of the Turkic languages.

In addition, we can see developments of areal grouping, during which Turkic languages of different genetic sub-branches form new territorial units and become involved in linguistic *areal interaction*—not only with each other but also with non-Turkic languages, leading to the develop-

[1] Here I would like to thank many colleagues whose works may not be mentioned in the bibliography, but who greatly inspired me to write this article. I can only beg their pardon if an article or a book of theirs contains special information and is not cited. This article is the product of some years of work, in which I may have learned facts but forgotten the sources from where I learned them. I want to extend special thanks to these colleagues, who spent time with me discussing the many different problems of language classification.

ment of new features within the Turkic languages in question. By a careful analysis of areal features one may discover linguistic connections of languages to areas to which they belonged in earlier periods of their development. Thus, at least some of the movements of the Turks can be reconstructed and compared with historical data. Paying attention to the genetic connections between the Turkic languages as well as to the areal ones makes the proposed classification model more complicated, but at the same time (as I hope) more appropriate for describing reality.

A common set of genetic features constitutes a *genetic string,* a set of areal features an *interactive area.* An interactive area may contain elements of different genetic strings. Vice versa, a unit belonging to a genetic string may also bear features of different interactive areas. Normally, isoglosses drawn by the features of one set are not totally congruent. Especially, with the growth of number of Turkic sub-units belonging to a super-unit, the number of features not attested in one or the other unit also increases. This leads to the fact that neither genetic strings nor interactive areas have sharp borderlines—they oscillate. If the analysis of linguistic data leads to diachronic conclusions, one should try to make use of historical data for external control.

I differentiate four main diachronic layers in the development of Turkic: New Turkic (TN), Middle Turkic (TM), Ancient Turkic (German *Alttürkisch,* TA) and proto-Turkic (pT). I use TA in the sense of pre-Chingisid Turkic, whereas the term Old Turkic (OT) is used for non-Islamic TA, i.e. Runic Turkic and Old Uigur. Then there follows the Middle Turkic period up to a date which is still open to investigation. Classical Middle Turkic ends in the 16th century. To set a border in time, I assume all products of spoken language from the second third of the 19th century on as New Turkic. The period between the 16th century and the beginning of New Turkic I consider a "twilight zone", for which no detailed information is available. For the written languages, one should keep in mind their high degree of conservativity, so that one must be aware of being confronted with monuments written at the beginning of the 20th century but, from a linguistic point of view, still belonging to Middle Turkic.

2. Common Turkic and Norm Turkic

From a practical point of view it seems necessary to define two statistic units: Common Turkic (ComT) and Norm Turkic (NormT). I call a feature Common Turkic, if it is attestable in all linguistic units of Turkic, or

if its loss in one, some or most of these units is explicable. A feature is called Norm Turkic, if it appears in a maximal group of linguistic units. Practically, the designation Norm Turkic covers languages like Turkish, Azeri, Tatar, Bashkir, Kazakh, Kirghiz, Uzbek, Turkmen and others, i.e. languages spoken by around 90% of contemporary Turkic speaking peoples, and especially Old Turkic. With the intention of designating these facts, I use the name "Norm Turkic" for the statistically most widespread type of Turkic, which is bound together by a whole set of common features. Like all features, the Common Turkic and Norm Turkic features may derive from genetic heritage or areal interaction.

2.1. Common Turkic

I can not present a catalogue of Common Turkic features here. To give some examples I only want to mention cases such as the verb **al-* 'to take', the basic numerals for the ones up to ten, or the preterite in **-DI*. As far as I know, these elements can be found in all Turkic units. At the present, there seems to be no monographic work on Common Turkic features.

2.2. Norm Turkic and non-Norm Turkic

As Norm Turkic (NormT) I designate such Turkic units which (i) show more or less complete loss of word-initial **h-* (see Doerfer 1981 and 1982), (ii) have kept the opposition **č- : y-* in word-initial position, (iii) show a nominal plural suffix **+lAr*, (iv) possess a gerund in *-B*, (v) display forms of the conditional suffix going back to **-sAr*, and (vi) have a 3rd person imperative suffix going back to **-ZUn*. By applying these features, three units can be designated as non-Norm Turkic (non-NormT): Chuvash, Lena Turkic (Yakut and Dolgan), and Khalaj.

2.2.1. Khalaj

As the only Turkic language, Khalaj has preserved quite consequently word-initial **h-* and shows a 3rd person imperative suffix *-tA*; the gerund in *-B* is missing in syntactically free use. Additionally we find a non-Norm Turkic locative in *-čA*. In the case of the plural and conditional suffixes, Khalaj behaves like Norm Turkic. As we know from Doerfer's investigations, it must have had long and intensive areal contacts with Oghuz (see 4.1.3, 4.1.3.1 and 4.1.3.3).

2.2.2. Chuvash

Chuvash is the only recent representative of the Bolgar Turkic subgroup and is most deviant from Norm Turkic. It shows replacement of the nominal plural suffix *+*lAr* by +sem, a suffix *-(t)tĂr* for the 3rd person of the imperative, which perhaps is to be connected with a causative form (see Erdal 1993) use of a gerundial unit *-sA* instead of *-B* (perhaps connected with the Old Turkic conditional suffix *-*sAr*; a *-sAr*-conditional is missing and seems functionally replaced by *-sAn*). By applying features of traditional classifications, it can be additionally individualized by other non-Norm Turkic, typically Bolgar Turkic features such as the sound changes *-*z(-)* > *-r(-)*, *-*d(-)* > *-r(-)* or *-*š(-)* > *-l(-)* or the use of ku instead of bo / bu as the demonstrative of proximity[2] **sarïġ* means 'white' and not 'yellow'. There exists a special possessive suffix *+Ăšĕ* of the 3rd person, which is only used with certain nouns (mainly kinship terms), numerals and pronouns (see Benzing 1959a: 736). The plural, possessive and case suffixes do not obey the regular Turkic order plural – possessive – case, but show possessive – plural – case. Chuvash (like earlier Volga Bolgar) has an ordinal suffix of its own different from the most widespread type *+*nč(I)* (see, e.g., Benzing 1954, Erdal 1993 and Adamović 1996). Chuvash has been subject to strong influences from Volga Finnic languages and from Russian. At the same time, it shows many correspondences with non-Bolgar Turkic and other phenomena that are also found in Mongolic, e.g. Turkic *-*d(-)* ↔ Chuvash, Mongolic *-r(-)*, *-*z(-)* ↔ *-r(-)*, *y- ↔ *ǰ-* ~ *č-*, *ti- > *či-* or lack of a pronominal *-n* in the nominative of the 1st and 2nd sg. pronouns.

2.2.3. Lena Turkic

In Lena Turkic (LenaT) the gerund in *-B* is replaced by *-An*, but the suffixes of the conditional and the 3rd p. sg. imperative *-TAr* and *-TIn* can be connected to the Norm Turkic forms of OT *-sAr* and *-zUn* by the sound change *(*z >)*s > t*, which is well attested in some frequent suffixes and some stems in Lena Turkic. Besides *+*lAr* we find additional plural markers such as *+t* or *+ttAr* < *+*t*+*LAr* (see Schönig 1988). Lena Turkic shows numerous additional specialities. As the only Turkic

[2] A comparable pronoun *go(l), gu(l), ko(l), ku(l)* can be found in Yellow Uigur (see Tenišev 1976a: 74), but there it is used in the same function as *o(l)*.

language it has *köt-* for 'to fly', which is perhaps connected with OT *kötör-* 'to raise, to lift up', and a deviant word for 'mouth' (see Doerfer 1988: 174, 1965: 171–172 and Schönig 1988). OT *qïrqïn* '(slave-) girl' has only survived in Lena Turkic in the plural form *kïrgïttar* to *kï:s* 'girl'. The Yakut comitative suffix *+lI:n* can easily be connected with the Old Turkic form *+lXGXn* (see Schönig 1991). Nominal, possessive and pronominal declensions show many deviations from each other. The survival of a dative-locative category like in Old Turkic may be the result of Mongolic and Tungusic influence, the lack of the genitive is perhaps due to interaction with Tungusic.

2.2.4. The Lena Turkic-Chuvash connection

Lena Turkic and Chuvash show a common set of non-Norm Turkic features such as preservation of the Old Turkic low vowel of the second syllable of *olor-* 'to sit down; to sit' (Brahmi texts; see, e.g., TT VIII) in Yakut *olor-*, Chuvash *lar-* and low vowels in suffixes such as the causative suffix *-DWr-* (where the other Turkic units normally have high vowels) or loss of the word-initial opposition *č- : y-*. Furthermore, the Old Turkic verb *tašïq-* 'to go out' (Yakut *taγïs-*, Chuvash *tux-*; in most modern Turkic languages *čïq-)* has survived in forms much closer to the Old Turkic one than in any other Turkic language. It is still open to discussion whether these common features merely point to a relatively early separation of Bolgar Turkic and pre-Lena Turkic from the other Turkic branches or whether at least some of them indicate in addition closer genetic connection or areal interaction.

3. Central Turkic and Border Turkic

As Central Turkic (CT) I consider all Turkic languages showing the sound change **-d(-) > -y(-)*, i.e. one of the classical features expressed by the word *ayaq* 'foot'. All members of Central Turkic are Norm Turkic. The non-Central Turkic units I call Border Turkic (BT). Border Turkic consists of the non-Norm Turkic units and some Norm Turkic ones.

Other features of Central Turkic are forms of the personal interrogative pronoun reconstructable as **kim*, existence of the verbs **toġ-* for 'to give birth; to be born' and **ket-* 'to go away (from)' (see 3.2.2.1), replacement of the 1st p. pl. ending *-mIz* by *-K* in the *DI*-preterite and the

conditional, dominant use of **bütün* for 'all, whole',[3] existence of the Old Turkic suffix **+lIK* and the privative suffix **+sIz* (existing in Chuvash and Khalaj as well).

3.1. Genetic internal segmentation of Central Turkic

Central Turkic shows three main branches: Oghuz, Kipchak and South East Turkic (SET); they seem to be tied together internally mainly by genetic strings. These branches and their sub-branches can roughly be characterized by the "classical" keyword **taġlïġ* as follows: In Oghuz at least reflexes of **-ġ* different from *-y* are preserved at the border of velar first syllables; it has mostly become *zero* in non-first syllables (e.g. Turkish *dağlı).* In Kipchak it is represented as **tawlï*, i.e. as *zero* or *-w* after the end of first syllables. In the Kirghiz-Kipchak (Kirghiz and Altay Turkic) sub-branch there is a strong tendency to generalize **-w* in nonfirst syllable-final position (**to:lu:,* see Benzing 1959b). South East Turkic shows the tendency to shift syllable-final **-G*-sounds to *-K (*taglïq, *taqlïq).*

The development of **taġlïġ* reveals the intermediary position of the Uzbek group of units between Kipchak and South East Turkic: Literary Uzbek shows *tåġli,* which I consider to be a geneticly relevant premodern Kipchak form. Other features such as the neutralization of *i : ï* (see Johanson 1986a) or loss of the pronominal *n* (uzb. *atida,* Nuig. *etidä* instead of *atïnda* 'on its horse') connect it with South East Turkic on the level of areal interaction. In accordance with historical sources (e.g. the *Babur-name),* we can reconstruct through linguistic features the development of Uzbek from an immigrant Kipchak unit to a mixed Kipchak-South East Turkic one. Today South East Turkic is best represented by the New Uigur (NUigur) group of units. Unlike Uzbek, it has an ablative suffix *+DIn* instead of *+DAn.* The same form appears in Lower Chulym Turkic, to which it may have been brought by New Uigur ("Bukharian") traders in the 17th century (see Pritsak 1959a: 624). For **taġlïġ* see also 3.2 and 6. We shall return to Central Turkic somewhat later.

[3] Chuvash also has *pětěm.* Mainly in North East Turkic we find numerous variations and alternatives such as AltayT **bastïra,* YenT **tekši,* Sayan Turkic: Karagas **tödö.* In Tuvan, among many alternants, we find *büdün* (like in Lena Turkic) or *xamïġ* (also attested in Old Turkic and Mongolic).

3.2. Border Turkic

The feature **-d(-)* has undergone various developments in Border Turkic. Its non-Norm Turkic constituents show *r* (Chuvash), *d* (Khalaj) and *t* (Lena Turkic). Its Norm Turkic constituents are Sayan Turkic (SayanT:**-d(-) > -d(-))*, Chulym Turkic (ChulT), Yenisey Turkic (YenT), Yellow Uigur (YUigur) and Fu-yü Kirghiz (**-d(-) > -z(-))*.[4] The latter three units seem to belong closer together, especially Yenisey Turkic and Fu-yü. Yellow Uigur in addition shows similarities to Sayan Turkic, particularly to Tuvan.[5] *z*-Turkic and *d*-Turkic (Sayan Turkic and Khalaj) have preserved final Old Turkic *-G*-sounds (**taġlïġ)*.

Individual features of Sayan Turkic are, e.g., a gerund in **-BIšA:n*, velar forms of the personal interrogative pronoun (see 3.2.1.1), loss of **ne(mä)* as the impersonal interrogative pronoun and ordinal suffixes such as Karagas (To"fa) *+š.KI (*? < **+nč+KI)*, and Tuvan *+KI*, which contain (or consist of) the old "correlational suffix" *+KI* (like in Old Turkic *il+ki* 'first'). Especially Tuvan has replaced the verb *čïq-* 'to go out, come out' (common to most Norm Turkic units but not to Khalaj) by *ün-*. Like Turkmen and Chuvash, Lena-Sayan Turkic does not use **-K* as a personal marker of the 1st p. pl. imperative. A *z*-Turkic (see fn. 4) speciality are *-ǰAŋ*-participles to express habituality, continuation, etc. Yellow Uigur has preserved the Old Turkic counting system (see fn. 8). Fu-yü has *ǰibir* for 'twenty' (see also 6.1). The most archaic forms of the Old Turkic word *älig* 'hand' have survived in Lena Turkic, Chuvash, Fu-yü and Yellow Uigur, where they still appear as bisyllabic, e.g. Yakut *ili:*, Chuvash *ală* (see also 4.1.1), Fu-yü *alïx*, YUigur *ïlïγ, elïγ, elig* (see Hu & Imart 1987; Tenišev 1976a and Doerfer 1989a: 186-187).

[4] In the strict sense of this definition we have to exclude the Shor-dialect of Khakas and the Mrass-dialect of Shor showing **-d(-) > *-y(-)* from Yenisey Turkic. A comparable situation can be found in Chulym Turkic, whose lower dialect shows *-y(-)*, whereas its middle dialect and Küärik have *-z(-)*.

[5] It is still unclear whether the Khalaj and the Sayan Turkic data point only to the relatively high degree of conservativity in these languages independently of each other or whether it is a sign of a common development in both languages going back to closer ties between them in ancient times. Another open question is whether LenaT *t* represents a totally independent development or whether it can be connected with the *d* in Khalaj and / or Sayan Turkic or the *z* of the other Border Turkic units (see Räsänen 1949: 29).

3.2.1. South Siberian Turkic

The Border Turkic units Yenisey Turkic and Sayan Turkic together with the transitory Chulym Turkic (see fn. 4) and the Kirghiz-Kipchak Altay Turkic (see 3.2.3) form the South Siberian Turkic (SST) area;[6] being a young areal group (see Schönig 1991), it has only a very few characteristic features common to every single unit within it, e.g. the use of the Old Turkic verb *yan-* 'to return, turn around', use of OT *bod* to derive reflexive pronouns (see 4.1.3.3) or loss of the plural marker *+lAr* in the 2nd person (see 6). For the word **kin(dük)* 'navel', see 3.2.4.2 and 4.1.3.2. One of the common features is the strong Samoyedic and Yeniseyic (Ket, Kot etc.) substrate already mentioned in Castrén (1857; see also Menges 1955-56 and Janhunen 1989).

3.2.1.1. South Siberian Turkic and non-Norm Turkic

Together with the non-Norm Turkic units Chuvash and Khalaj most South Siberian Turkic units have forms of the personal interrogative pronoun which are *not* reconstructable as CT **kim*. Chuvash, Khalaj, Yenisey and Altay Turkic show forms reconstructable as **käm*. It is impossible to decide whether Lena Turkic belongs to the **käm-* or the **kim*-group. Sayan Turkic together with Fu-yü gain a special profile by showing forms such as Tuvan *qïm* or Karagas *qum*, Fu-yü *gïm*.

[6] South Siberian Turkic includes the literary languages and the units, which in Soviet literature are normally treated as their dialects, of Altay Turkic (formerly called "Oyrot"), Yenisey Turkic (lit. languages Khakas and Shor) and Sayan Turkic (lit. languages Tuvan and Karagas / To"fa); Chulym Turkic has never developed a literary language and remains between Altay Turkic and Yenisey Turkic. I prefer the geographical designations to avoid too many associations with the designations of the mostly synthetic literary languages, which sometimes do not obey the isoglosses drawn by features of the units belonging (as "dialects") to the single groups (see fn. 4). Units such as (eastern) Siberian Tatar, Yellow Uigur and Fu-yü behave in some respects very much like SST. As mentioned by Doerfer 1985: 1, (eastern) Siberian Tatar is, in many repects, a transitional group between SST and what I call Volga-Ural-Caucasus Kipchak Turkic.

3.2.2. North East Turkic

South Siberian Turkic together with Lena Turkic formed another interactive areal not too long ago—North East Turkic (NET). More or less exclusive features of North East Turkic are the *nomen actoris* in *+A(:)ččI* and the total replacement of the privative suffix *+*sIz* by constructions of the type *noun (+ possessive suffix)* + **yoq*, which can be found beside *+*sIz* in other Turkic languages as well (Schönig 1991). As the North East Turkic protoform of the numeral for 'thousand' we can reconstruct *muŋ* (see also 3.2.4.2 and 6.1). As in Karachay-Balkar, Fu-yü, Salar and Yellow Uigur a replacement of those Old Turkic tens, whose connection with the corresponding ones is not transparent, has taken place in analogy to OT *sekiz on* 'eighty' and *toq(q)uz on* 'ninety'.[7] In South Siberian Turkic this replacement is perhaps inspired by transparent derivations of tens from ones in neighboring or substrate languages (Mongolic, Yeniseyic, Samoyedic, etc.); the total loss of *+*sIz* and the form of the *nomen actoris* are surely due to Mongolic influences.[8] The same holds true for the reflexive pronouns, where the special

[7] As in the case of phonotactic rule sets (see 5), units belonging to different subbranches show the same development. Karagas and Yellow Uigur start with 'twenty', Tuvan with 'thirty', Fu-yü, Altay Turkic and Lena Turkic with 'forty'. Only in the YenT-ChulT-group starting with 'sixty', is the distribution in accordance with subdivisions known from other features. In Castrén's Karagas, even the word for 'thousand' is decimalized (*on ~~d~~üs)*. Salar has an alternative additive system starting with 'sixty' *elli on* ~ **altmiš* (see Tenišev 1976b: 121). In Karachay-Balkar we find a vigesimal system inspired from neighboring Caucasian languages (see Pritsak 1959b; JN: 220).

[8] That the formally incoherent Old Turkic system of ones and tens really is old is proved by the fact that Yellow Uigur, which shows a very strong tendency to decimalization, uses in its archaic counting system *üčon* for 'thirty', but *pes otïs* for 'twenty-five'. The privative **yoq*-constructions could be inspired by Mongolic patterns with **ügei* ('does not exist' = **yoq)*, e.g. Kalmyk *üzü:r* 'Ende, Gipfel etc.', *üzü:r ugE:* 'end-los' (KlmWB 460b). For the *nomen actoris* in *-A(:)ččI* we find the Mongolic *nomen imperfecti*, Classical Mongolian *-Γa*, Buryat *-A:*, Khalkha *-A:* (see Poppe 1955: 273; Sandžeev 1964: 136-137) in connection with the Turkic-Mongolic suffix *+čI* "to form nouns of vocation" like in *alaġači* 'killer' < *ala-* 'to kill' (Poppe 1954: 45). Even the replacement of *toġ-* by *törö-* (s. 3.2.3) belongs here. The Mongolic verb for 'to be born' is *törö-* and has strengthened the homonymic Turkic verb against **toġ-*.

position of Lena Turkic within North East Turkic can easily be demonstrated. Lena Turkic uses *bäyä*, a borrowing from Mongolic, where it originally ment 'shape, body'—like OT **bod*, which is used as the base of the reflexive pronouns in South Siberian Turkic (see also 4.1.3.3). But while South Siberian Turkic has only copied the Mongolic model, Lena Turkic has made a material copy. Furthermore, Lena Turkic has a 3rd p. sg. pronoun not identical with the demonstrative pronoun of distance (see Schönig 1995d). This is due to the fact that Lena Turkic is part of a young North Eastern Siberian interactive area including mainly Tungusic and Mongolic languages, from which it has received some strong non-Turkic impulses (see Schönig 1988 and 1993d).

There remains a whole set of features which is common to most North East Turkic and South Siberian Turkic units, but is not attestable in Altay Turkic (see 3.2.3).

3.2.2.1. North East Turkic and Chuvash

The Old Turkic verb *ï:d-* 'to send' is kept as a full verb in Lena Turkic *ï:t-*. South Siberian Turkic and Chuvash have also preserved the "short form". But Chuvash *yăr-*, Altay Turkic *ïy-*, Khakas *ïs-* and Tuvan *ït-* also function as auxiliary verbs; in Yenisey Turkic and Sayan Turkic, combinations of this verb and the gerund in *-B* have gained suffixal status such as Khakas *-(I)vIS-* or Tuvan *-(I)vIT-*. "Short forms" have survived in the Kipchak languages Karaim, Kazakh and Kirghiz as well (see also 4.1.1.2).[9] North East Turkic, like Chuvash, does not show the verb **ket-* 'to go away (from)'. While Chuvash uses the verb *kay-*, North East Turkic has lost it in most of the units without a direct or simple lexical substitute. Both groups show a tendency towards desonorization of word-initial and word-final obstruents and sonorization of intervocalic ones.

The connection between Chuvash and North East Turkic goes back to old areal language contacts. There must have been Bolgar Turkic groups

[9] But they also show forms going back to the combination **ï:du ber-*, as Kipchak languages and the South East Turkic New Uigur normally do. According to dictionaries and grammars none of the forms have auxiliary functions in Karaim, in Kazakh only the long form is used as an auxiliary, in Kirghiz both forms. Kirghiz has a third form *ir-* functioning only as an auxiliary verb (KrgRS 302b). The question arises whether this *ir-* is a Bolgar type development of **ï:δ-* (cf. Chuvash *yăr-* id.) (see Schönig 1991 and 1995b).

in Southern Siberia, maybe as early as at the time of the Hsiung-nu confederation (see Róna-Tas 1980 and Janhunen 1989: 294). Besides loanwords, one important argument of Róna-Tas (1982) for old Bolgar-Siberian Turkic connections is the shape of the word for 'stirrup'. One of the groups shows forms pointing back to a protoform with labial initial vowel and one group—North East Turkic, Chuvash and Yellow Uigur—with illabial initial vowel. Only Altay Turkic has a labial vowel and does not behave like a North East Turkic language.[10]

3.2.3. Altay Turkic

Altay Turkic often behaves ambiguously. On the one hand, it shows the Central Turkic feature **-d(-) > -y(-)* and has not, for example, lost the frequent Turkic derivational suffix **+lIK*, on the other hand it has numerous non-Central Turkic features common to South Siberian Turkic and North East Turkic. Thus it has, differently from remaining North East Turkic, Chuvash and Khalaj, not kept postvocalic **-yUr*-aorists

[10] See Altay Turkic *üzeŋi* (RAltS 738b), Khakas *ĭzeŋe* (RXksS 834a), Tuvan *ezeŋgi* (TuvRS 577b), Karagas *ez'eŋge* (Rassadin 1971: 183), Yakut *iŋehe* (RJakS 623a), *iŋahe* (Rassadin 1971: 183), YUigur *ezenky* (Tenišev 1976a: 179). In Erdal (1993: 161) we find the following comment on the word for 'stirrup': "Nun kann auch aus einem frühen altuigurischen Text die Form *izäŋü+lük* hinzugefügt werden; siehe Erdal 1991: 128-9. Sie zeigt, daß vermutlich der *i*-Anlaut der ältere ist und die Form mit /ü/ durch einen volksetymologischen Abgleich an *üzä* entstanden ist. Für den Auslaut bleibt die schon genannte Isoglosse /U/ gegen bolgarisch und sibirisch /A/. In diesem Fall wie auch in allem anderen stimmen die im Mongolischen zu findenden Elemente mit der bolgarisch-sibirischen Gruppe überein." Erdal may be right in his assumption, but it should at least be argued for, because there is no general "Siberian" form of this word showing final /A/. As we can see from the data given above, Altay Turkic and Tuvan show high final vowels. One could argue that the Altay Turkic data are of no importance here because it again does not behave like SST. The Tuvan (and perhaps the Yellow Uigur) form may be explained as a metathetic form (Tuvan *ezeŋgi* < *izeŋge)*. The Yakut forms have a metathetic protoform as well (*iŋehe* < **iŋese* < **iseŋe; iŋahe*, if it is not a misprint, shows velarization of the vowel neighboring *ŋ*). But it is impossible to say whether it really goes back to a form with a low final vowel because in Lena Turkic high word-final vowels can be lowered. However, the parallel Karagas form may strengthen this assumption.

instead of *-*r* like most Turkic units,[11] and consequently no longer shows a formal correspondence between the postvocalic forms of the vocalic gerund and the aorist *-yU – -yUr*.[12] Some features common to Chuvash and most North East Turkic units are absent in Altay Turkic as well. We find no consistent preservation of the Old Turkic 1st p. pl. ending *+mIz* in the *di*-preterite nor replacement of the verb **toġ-* 'to give birth; to be born';[13] in North East Turkic units it is normally replaced by *törö-*, while Chuvash has *śura-* 'to give birth' *(< *yara-)*. Moreover, Altay Turkic shows typical Kirghiz-Kipchak developments of sound groups consisting of a palatal labial vowel and *-g* or *-ŋ* (see 4).

From the lack of some North East Turkic or South Siberian Turkic features we can deduce that Altay Turkic did not have intensive and / or longlasting contacts with the other members of these groups.

3.2.4. Other connections between Lena Turkic and single South Siberian Turkic units

There are some remarkable common features between Lena Turkic and single South Siberian Turkic units, which perhaps can be assumed to be sporadically preserved products of closer internal contacts during the phase of formation of North East Turkic. As one can guess from the geographical distribution of the North East Turkic constituents, such common features can be found most frequently in Lena Turkic and Sayan Turkic, especially Karagas.

Together with Altay Turkic (and Fu-yü) Lena Turkic substitutes the original tens beginning with 'forty' (see fn. 7) and shows the same distribution of intervocalic consonants in numerals (see 6.1). The Lena

[11] In Khalaj we find postvocalic *-yUr* (Doerfer 1980 and 1988), for Chuvash, see Johanson 1976a: 135-136. In Lena Turkic we find *-I:r*, in Khakas *-ir* (with neutral *i* pointing to a contraction). In Shor and Sayan Turkic, the quality of the stem-final vowels decides the quality of the contraction product.

[12] The Old Turkic connection between the suffixes of the vocalic gerund and the aorist *...C-V(r); V-yU(r)* is preserved in Lena Turkic and Yenisey Turkic in *...CA(r); ...I:(r)* or *...i(r)*. For Sayan Turkic, see Schönig (1989). Altay Turkic with its postvocalic *-*r*-aorist and *-*y*-gerund behaves like a Kipchak language.

[13] The ambiguous meaning of the verbs **toġ-* and **törö-* (transitive or intransitive) does not seem to follow genetic strings or areal patterns. For example, in North East Turkic it is intransitive in Yakut, Khakas and Karagas, but transitive in Altay Turkic and Tuvan.

Turkic *-IAx*-future could go back to **-GO oq* (the Old Turkic *nomen futuri* + particle *oq)*, which in Chulym Turkic is preserved as *-GOK*. The Lena Turkic suffix of the comparative case *+TA:GAr* resembles the Khakas suffix *+DAŋAr*, which has a comparable function (see XksGr. 267-268 and Borgojakov 1976). Poppe (1959: 681) supposes that it is taken from Mongolic. In both branches, the interrogative pronoun **qa:ńo täg* 'which type, how' has developed into **qay*-stems: in Lena Turkic (e.g. Yakut *xaydak*) and Yenisey Turkic (e.g. Shor *qaydi, qaydïg,* Khakas *xaydi, xaydag)*. In the other Turkic units we mostly find **qan*-forms.

3.2.4.1. The Lena-Sayan Turkic area

The Lena-Sayan Turkic area can be defined by the absence of some widespread Central Turkic features (most of which appear in Chuvash and Khalaj as well), e.g of the verb **säv-* 'to love' (which seems to be absent in Khakas), of the nonpersonal interrogative pronoun going back to **ne(mä)* or of verbal combinations with **bašla-* to express 'to begin to'. Moreover, Lena-Sayan Turkic uses **qïl-* but not **et-* for analytic denominal derivations of verbs (see also 3.2.4.2 and 3.2.5). It shows a very low level of formal recursivation (see also 4.1.1.2). Lena Turkic has no stable cursivity markers (see Buder 1989). Sayan Turkic has cursive participles for anteriority and nonanteriority using the "classical" set of the four auxiliary verbs expressing nontransformativity *tur-* 'to stand (up)', *olur-* 'to sit (down)', *yat-* 'to lie down, to lie', *yür(ü)-* / *yor(ï)-* 'to go' as markers.[14]

On the other hand, Lena-Sayan Turkic shares some non-Central Turkic features with Chuvash such as the absence of **-nčI*-ordinals (see also 2.2.2)[15] or the quite frequent sound change **a > ï* in first syllables, which may have a common source in the two different branches. Together with Turkmen, Lena-Sayan Turkic shows suffixes of the inclusive and exclusive 1st plural persons in the imperative paradigm with the

[14] It is impossible to reconstruct a common pronounceable form for a verb of the type **yV^{lab}r(V)-* of this meaning.

[15] The Lena Turkic suffix *+(I)s* may be reconstructed as **+nč* (see Schönig 1991; for Sayan Turkic see 3.2). According to Thomsen 1959: 566 (mainly based on Malov's works), Yellow Uigur has an ordinal suffix *+nč,* too. But Tenišev (1976a: 74) only mentions a form such as *+(I)ndzï* or *+(I)ndzïlïq*.

common structure *1st p. inclusive = 1st p. exclusive + 2nd p. pl.*[16] For the forms in Kipchakoid South Siberian Turkic and some Bashkir dialects, see 3.2.5. Nasality of OT *ń* (palatal *n*), which is also preserved in Khalaj *(-n(-))* and Oghuz *(-yVn(-))*, has partly survived as *"nasal y"*.[17]

3.2.4.2. The Lena Turkic-Karagas connection

Within Sayan Turkic, Karagas shows even closer connections to Lena Turkic by features only attestable in these two units, e.g. loss of the numeral for 'thousand' and its replacement by a Russian loanword[18] or a verb **qïn-* (Yakut *gïn-*, Karagas *qïn-)*, which is mainly used to derive onomatopoeic verbs (see Ubrjatova 1985: 149; JakGr. 225 and Rassadin 1978: 155-157). For 'navel' only the short form of **ki(:)n* exists (see 4.1.3.2). Both units have a partitive case, expressed by suffixes formally identical with the Old Turkic locative-ablative suffix *+DA*. Furthermore, we may assume that the Old Turkic ordinal suffix *+nč* has only survived here (see 3.2). As the only North East Turkic units, Karagas and Lena

[16] I have earned a good deal of criticism for assuming the *inclusive : exclusive* opposition in the 1st p. pl. of the imperative in Turkmen and North East Turkic as an archaic feature only reconstructable from New Turkic data. I do not think that this point is very important in our framework, but I want to point to some facts. We find this opposition mainly in languages showing a high degree of conservativity. Maybe the existence of the categories in North East Turkic may be explained by areal influences from Mongolic or other neighboring languages. But why then do they appear only in the imperative and not, e.g., on personal pronouns like in these neighboring languages? If these categories have come into existence "spontaneously", why then only in the imperative mainly in border languages and in branches which definitely had no language contacts for hundreds of years? But as I said before, the point is of no real importance for this classification model, it would only help to strengthen a set of features already existing between the border languages.

[17] For Dolgan, see Ubrjatova (1985: 38), for Yakut, see JakGr. (61-62), for Sayan Turkic, see Menges (1959b: 652), especially for Karagas, see Rassadin (1971: 49). For Oghuz and Khalaj, see Doerfer (1971: 178).

[18] We find Dolgan *tï:haččа,* Yakut *tïhï:nča* and Karagas *tï:sičči*. The original form *muŋ* can be reconstructed for Yakut by means of instances such as Radloff (1908: 40) *muŋ* '100 Rubel' and JakRS (244b) *muŋ* 'border, ultimate degree'; *muŋ älbäx* 'very much'. BöWB (150a): in the 19th century *muŋ* was already "ein altes, nicht mehr gebräuchliches Wort".

Turkic show **qač-* besides the **qanča*-form of the interrogative 'how much, how many' (see 6).

3.2.5. Kipchakoid South Siberian Turkic

Altay and Yenisey Turkic together with Chulym Turkic sometimes demonstrate features different from Sayan Turkic, but common to Kipchak. The division of South Siberian Turkic into Kipchakoid South Siberian Turkic and a Sayan Turkic branch can be seen, for example, in the distribution of *K* as a 1st p. pl. personal marker, which only appears in Kipchakoid South Siberian Turkic (mainly in the imperative). It is hereby more closely connected mainly to the modern Kipchak languages, to Azeri and Lena Turkic, which also use **K* as a personal marker in the imperative paradigm.[19] Other features separating Kipchakoid South Siberian from Sayan Turkic are the form of the suffix of the 1st p. pl. inclusive—the structure *1st p. inclusive = 1st p. exclusive + (plural marker)*—like in some Bashkir dialects, the use of *et-* as an auxiliary verb for denominal verb derivation like in Oghuz, Kipchak and South East Turkic and marking strategies in the participal systems of Kipchak and South East Turkic. By analogization of case-suffix-final nasal consonants and the distribution of velarity / palatality of the numeral 'twenty' it is more closely connected to Kirghiz, which, at the same time, is separated by these features from the remaining Kipchak group (see 5.1.1 and 6.1).

I call this group Kipchakoid instead of Kipchak because an important feature of modern Kipchak, the preservation of intervocalic **-t-*, has been given up and the forms of the numerals with intervocalic consonants have stabilized in forms more or less different from the current Kipchak type (see 6.1). I assume these features to go back to perhaps several (proto-) Kipchak groups involved on several steps in the glotto-

[19] I reconstruct the Lena Turkic exclusive and inclusive forms *-IAx* and *-IAγlŋ* as **-AyIK* and **-AyIK + Iŋ*, i.e. going back to the suffix *-AyIK*, well-known in most of the modern Kipchak languages. Even if this is not correct, the fact remains that *K* is a typical sign of Kipchak imperative paradigms from the time of the Codex Cumanicus on, which shows *-AlIK*. This form appears in modern Kirghiz and is the protoform of the exclusive imperative forms in Altay Turkic dialects, in Chulym Turkic and in Yenisey Turkic Kyzyl. The Azeri form may have emerged by internal analogizations or by areal contacts with Kipchak tribes (see Schönig 1987b).

genesis of South Siberian Turkic and North East Turkic. Perhaps some Kipchak(oid) layer is responsible for the representation *čač of the Old Turkic word for 'hair' *sač* (see 5).

(To be continued.)