

The Turkish Y and Related Sounds

By Sir GERARD CLAUSON (London)

Terminology. I use "Turkish", not in the limited meaning of 'the language of Turkey', but as a generic term for all the languages genetically related to the language of the Türkü Dynasty (6th to 8th Cent. A.D.), from whose name the word is of course derived, including those anterior to that date. In other words I use "Turkish" where some other scholars use "Turkic", a word which seems to me open to the objection that if the Greek adjective *tourkikos* is to be used in English it must be transcribed consistently either as "Turcic" or "Turkik", both of which look grotesque. Similarly I use "Mongolian" as a generic term for all languages genetically related to the language of the 13th Century Mongol-un Ni'uca Tobçi'an, or "Secret History". I use dashes to avoid constant repetition of the words "initial", "medial" and "final"; thus *y-*, *-y-* and *-y* respectively mean initial, medial and final *y*.

Transliteration. I use the official alphabet of the Turkish Republic with 7 additional letters for sounds not represented by that alphabet: (1) *đ* for the voiced dental spirant in English "the"; (2) *é* for the closed *é* in French *été*; (3) *ñ* for the palatal nasal in Spanish *Señor*; (4) *ŋ* for the guttural nasal in English *sing*; (5) *t'* for the palatalized *t* in Russian *telo* (pronounced *tyelo*); (6) *w* for the English *w*; (7) *x* for the unvoiced guttural fricative in German *machen* or Scottish *loch*. On one minor point I depart from the usage of the Turkish Official Alphabet: I use *ğ* only with back vowels, and to represent the voiced guttural fricative, the sound of the Arabic *ğayn*; this seems to have been the way in which this letter was pronounced in early Turkish. I distinguish long vowels, which existed in early Turkish, by a colon, e. g. *a :* for long *a*; the ordinary superimposed dash is too clumsy over *ö* and *ü*. For the sake of uniformity I use the same alphabet for transcribing Arabic with the addition of ' for *ayn*, *q* for *qa :* *f* and subscribed dots under "hard" *h*, *đ*, *s*, *t* and *z*.

Dating. I use Roman numerals, designating the centuries A.D., to date individual words, texts and languages, in four formulae: (1) VIII marks a word etc. known to have been in use between A.D. 701 and 800; (2) VIII/LX marks a word etc. in use between A.D. 701 and 900, but whether before or after A. D. 800, or both, is uncertain; (3) XVff. marks a word etc. used after A.D. 1400 for an undefinable period of more than a century; (4) pre-XX marks a word etc. used for an undefinable period of more than a century prior to A.D. 1901.

It is a real pleasure, as well as a privilege, for me to contribute to a birthday gift to my friend Prof. N. N. POPPE, and I do not think that I could find a more appropriate topic on which to address him than one which really arises from our long talks together in London in the summer of 1956.

A great deal of work has been done on Turkish phonetics during the last hundred years, but nearly all these studies have been, so to speak, horizontal; they have been directed to presenting the complete phonetic structure of a single Turkish language or group of languages, or a conspectus of the forms taken by one or two significant Turkish words in a wide range of languages, at a single point in time, usually the present day. It seems to me that it would be useful, for a change,

to approach the subject, so to speak, vertically, that is, to try to work out the whole life history of one or two selected Turkish sounds during the last 1200 years, and then to see whether there is any means of projecting that history further back than the first half of VIII, the earliest period for which we have substantial Turkish texts.

To avoid any possible misunderstanding, I should explain that this study is based almost entirely on a comprehensive analysis of the occurrence at all periods and in a wide range of languages of all known Turkish words beginning with vowels and most of those beginning with *c-* and *ç-*; outside these limits I do not claim to possess adequately coordinated information or to speak with authority; even within them I cannot guarantee that there are no errors or omissions; indeed, if precedent is a guide, there probably are some of both.

The earliest history of Turkish is wrapped in mystery, and indeed the question whether it is genetically related to Mongolian (in future called "Mong.") and the Tungus languages is still hotly debated. My own view, for what it is worth, is that it is not, and I have stated my reasons in an article published in the *Central Asiatic Journal* II 3. Accordingly I have written this article on the assumption that, if the same word occurs both in Turkish and in Mong., it is native to one and a loan-word in the other, and that a Turkish loan word in Mong. may in appropriate cases be used to elucidate the life history of that word in Turkish.

First I must summarize briefly the Turkish languages included in, and excluded from, this study. I accept the view expressed in Dr. PRITSAK's book *Die Bulgarische Fürstenliste und die Sprache der Protobulgaren*, Wiesbaden, 1955 (in future called "Pritsak, 1955") that the Huns spoke an early Turkish of some kind, but our information about Hunnish is so dubious and obscure that I leave it out of account. I accept, too, his theory that Proto-Bulgar is a later form of Hunnish, or at any rate the language of some closely related Turkish tribe that formed part of the Hunnish movement to the West, and that Chuvash is a later form of Proto-Bulgar, but Chuvash has departed so far from the main stream of Turkish that I leave it too out of account. Similarly I also accept provisionally A. P. OKLADNIKOV's theory, expounded in *Istoriya Yakutskoy A.S.S.R.*, Moscow-Leningrad, 1955, that Yakut is a later form of the language of the Uç Kurikan, mentioned in the VIII Türkü inscriptions, but Yakut has departed so far from the main stream of Turkish that I leave it too out of account.

It is still an open question whether the "Runic" inscriptions in the valleys of the Yenisey and Talas are the earliest actual Turkish texts (as opposed to isolated words and phrases in foreign authorities). Our Russian colleagues claim that they go very far back, perhaps to VI, and this may be right; but it is equally arguable, and I think more probable, that they are not so much archaic, as late and provincial. In any event they are mostly incoherent and fragmentary, and I have left them too out of account.

The earliest Turkish texts which can be firmly dated are the "Runic" inscriptions found in the Orkhon valley and associated with the Northern Türkü Dynasty. The language of these texts I quote as *VIII Türkü*. The language of the paper documents in the same script (the *Irak Bitig* and minor texts) I quote as *VIII/IX Türkü*.

There are two inscriptions in the same script associated with the first Uyğur Dynasty, those at Şine-usu and Kara Balğasun (the latter mere fragments); I quote the language of these as *VIII Uyğ.*

The name "Uyğur" has been much misused. The Chinese historians tell us that the Uyğur did not emerge from obscurity till the beginning of VIII, and there is no reasonable doubt that references in Chinese histories anterior to that date to peoples whose names have hitherto been transcribed as "Uyğur" are in fact references to the Oğuz. Most of the standard works on the early history of the Turks are bedevilled by references to "Uyğur" who are really Oğuz, and some, for example WIEGER's *Textes Historiques*, by references to "Tungus" who are merely *Tung Hu* 'Eastern barbarians.' In the last century "Uyğur" was the term used for any Turkish language written in Uyğur script or the Mongolian Official Alphabet derived from it, a practice hardly more sensible than would be one of calling every language written in the Latin alphabet "Latin". It has now been confined within more sensible limits and such texts as the *Kutadğu Bilig* are no longer said to be in Uyğur, but our Russian and Chinese colleagues have recently extended it again by calling the modern Turkish of Sinkiang "Neo-Uyğur", a name for which there does not seem to be much real scientific basis.

Obviously Uyğur is the right term for the inscriptions of the first Uyğur Dynasty and there is contemporary authority for using it for the language of Turkish Buddhism. Texts in this language were still being copied in Buddhist monasteries in XVIII but by that time it was probably as much a dead language as Latin is today. The real problem is whether the term can properly be applied to the contracts etc. in Uyğur script, one firmly dated as not earlier than Tuğluk Temür (A. D. 1347-1363), such as those published in RADLOV's *Uigurische Sprachdenkmäler*. Other documents, still unpublished, of the same kind are associated with the Karakhanid Dynasty, the official language of which was Xakani (see below), but pending further elucidation I still call the language of these documents Uyğur.

In the Turkish translations of Manichaean scriptures etc., for reasons too complicated to explain here, I distinguish provisionally three dialects quoted as *VIII/IX Man. A-dial.*, *Man. Ñ-dial.* and *Man. Uyğ.* The translations of Buddhist texts I quote as *IX ff. Bud. Uyğ.*; some may well be a little earlier and some are certainly later, IX ff. seems a fair compromise. The Uyğur documents other than these, medical texts, contracts etc., I quote as *IX ff. Civ.* (i.e. Civil) *Uyğ.* for the medical texts and *XI/XIV Civ. Uyğ.* for the rest, unless they can be dated more precisely.

The language of the Turkish words in Mahmud al-Kaşğari's *Di:wa:nu'l-Luğa:ti'l-Turk* and of the *Kutadğu Bilig*, which was the language current in the domains of the Karakhanids and is conveniently called "Xakani", I quote as *XI Xak.*; the specific references to the former, e. g. Kaş. I 32, are to B. Atalay's Turkish translation. There are one or two Xak. texts like the *Atabatu'l-Haqa: yiq* which are later than XI, perhaps even XIII, but they are of minor importance.

Material later than this I divide into six geographical groups. This seems to me the minimum number which makes scientific classification possible. Apart from a very few marginal cases, the languages within each group have closely related vocabularies and phonetic peculiarities, while there are quite significant differences between the vocabularies and phonetic peculiarities of languages in different groups.

The groups are the following:—

NE: The Turkish languages of Eastern Siberia (except Yakut and Tarançi), for which the principal and, apart from CASTRÉN etc., the earliest authority is the XIX entries in RADLOV's *Wörterbuch* of words in the dialects of the Altai, Kaça,

Kızıl, Koibal, Kumandu, Küerik, Lebed, Sagay, Şor, Teleut, Tuba, and what he called Soyon, but I, following modern practice, call Tuvan. I use abbreviations based on Radlov's for these languages. I have also used for XIX Tuvan N. F. KATANOV's *Opyt Isslédovaniya Uryankhayskavo Yazyka*, Kazan, 1903. For NE XX I have used N. A. BASKAKOV's *Khakassko-Russkiy Slovar'*, Moscow 1953 (*Xakas*) and A. A. PALMBAKH's *Tuvinsko-Russkiy Slovar'*, Moscow 1955 (*Tuv.*).

SE: The written language of Chinese Turkestan, which, following the traditional practice, I call *pre-XX Türki*, is no doubt directly descended not from Uyğ. but from Xak. This must have been the language spoken in this area when the Karahanids ruled Kashgar and Yarkend, but there is a long gap in the tradition, and the earliest specimen that I know of *pre-XX Türki* (there may well be earlier ones) is the XVIII Turkish column in Chien Lung's "Five Language Mirror" in the British Museum, which I have not so far used. Other languages in the group are XIX Tarançi (*Tar.*), quoted in RADLOV's *Wörterbuch* and several spoken dialects recorded by various scholars.

NC: The earliest references to languages in this group are the XIX entries of Kazax (*Kzk.*), Kırğız (*Kır.*) and Tobol (*Tob.*) words in RADLOV's *Wörterbuch*. Radlov unfortunately calls Kazax "Kırğız" and Kırğız "Kara Kırğız", which is liable to cause some confusion. For *XX Kır.* I have used Taymas' Turkish translation of Yudahin's dictionary and for *XX Kzk.* MUSABAEV's *Kazakhsko-Russkiy Slovar'*, Alma Ata, 1954.

SC: Late Xak. merges imperceptibly into the earliest language in this group, but as the line must be drawn somewhere I classify as SC XIV the Turkish vocabulary material in the *Hilyatu'l-insa:n wa ḥalbanu'l-lisa:n* of Cama:lu'l-Di:n ibn Muhanna (*Muh.*) and the language of Rabḡuzi's *Qisasu'l-Anbiya:* (*Rbḡ.*) and the Nahcu'l-Fara:dis (*Nahc.*). I classify the classical language of Nawa:'i and his followers as *XV ff. Çağ.*, preferring the traditional, though technically incorrect, name Çağatay to the new Russian term "Old Uzbek," which suggests that it was originally the language of the Özbegs, whereas they found it already well established when they first became prominent. Later stages of the language I call *XIX* and *XX Özb.*

NW: The Kıpçak (*Kıp.*) group, starting with the references to Kıpçak in Kaş. and continuing through the XIII–XV Kıpçak-Arabic vocabularies and the XIV Codex Cumanicus (*Kom.*). Later languages of the group are the Karaim of Krim, Lutzk and Troki (*pre-XX Kar. K., L. and T.*) recorded in RADLOV's *Wörterbuch* (and Kar. T. in KOWALSKI's monograph) and the XIX dialects of Kazan (*Kaz.*) and Krim (*Krm.*) in the same authority. Other languages in the group include Karakalpak (*Kk.*), the authority for which is N. A. BASKAKOV's *Karakalpakskiy Yazyk*, Moscow 1951, and Bashkir.

SW: The Oğuz (*Oğuz*) group starting with the references to "Çużz" in Kaş. and continuing through the earliest texts from Anatolia (*XIII etc. Anat.*) and Osmanli (*pre-XX Osm.*) down to the present day language of Turkey, which can conveniently be called "Republican Turkish" (*XX Rep.*) and the spoken dialects of Anatolia (*XX Anat.*), together with related languages like Azeri and Türkmen.

It is through these languages, extending from VIII to XX that I propose to trace the history of the Turkish sound *y* and part of the history of the other sounds which at some date turned into *y*, or into which at some date *y* itself turned.

I must make one preliminary reservation. Languages are not scientific constructions, they are means of communication used by human beings, who in order to make themselves understood instinctively follow certain conventions, but do not consciously conform to immutable phonetic laws. What I am trying to do is to reconstruct the history of certain sounds, but there must always have been some words, like *ev*, *idi*: and *adut* mentioned below, which obstinately took a line of their own.

I will start by working backwards and consider the earlier history of XX *y* and, in the first instance, *-y/-y*. It is of course common knowledge that this sound quite regularly represents four different VIII sounds:— *-y/-y*, *-d/-d*, *-ñ/-ñ* and certain velar and guttural sounds.

(1) *-y/-y* < *-y/-y*. Examples are *ayak* “cup, bowl”, which occurs, in this form only, in VIII/IX Türkü, XI–XIV Civ. Uyğ., XI Xak. and all geographical groups and as a loan word in the form *ayaxa*¹ in XIII Mong., and *a:y* “moon, month”, which occurs, in this form only, in every Turkish language from VIII/IX Türkü onwards.

(2) *-y/-y* < *-d/-d*. Examples are VIII *adak* “leg, foot” which appears as *ayak* in most XX languages and *ö:d* “time” which appears as *öy* in several later languages. As these words appear in other forms in other languages they will be discussed later in the section dealing with the evolution of *-d/-d*.

(3) *-y/-y* < *-ñ/-ñ*. *Ñ* was always a rare sound; it occurs in its pure form in VIII and VIII/IX Türkü and as *-n/-ny-* in VIII/IX Man. *Ñ*-dial. . Kaş. (I 31) says that “the people of Arğu (defined by him as the area between Balasagun and Taraz, that is due west of the Issik Köl) pronounce all their medial and final *ya*:’s as *nu:n’s*”, giving three examples which will be discussed later. This is no doubt a mistake; *ñ* occurred only in a limited number of words, and presumably in these words tended to turn into *n* rather than *y* in Arğu. Other words in which the sound occurs are VIII Türkü *yañ-* “to scatter, disperse”, to be carefully distinguished from *ya:d-* “to spread” (both later became *yay-* and are indistinguishable in later languages) and the proper name spelt *Kitañ* in VIII Türkü (a form confirmed by early Chinese authorities) and *Kıtay* everywhere else.

(4) *-y/-y* < velars and gutturals. This sound change occurred in two quite different areas. It must have occurred very early in the NE languages, since it can be observed in some of the earliest Turkish loan words in Mong., e. g. *takiya* ‘a hen’ from Turkish *takiğu* and *ayıl* ‘cattle-pen’ from Turkish *ağıl* (it is uncertain whether forms like NE XIX Tel. *ayıl* are survivals or reborrowings from Mong., probably the latter). In this area it occurred both with back and with front vowels; the change must have been from a voiced *ğ/g*, but in some words these were transitional sounds from an earlier unvoiced *k*; for example VIII Türkü *éki*: ‘two’ became *igi* in NE XIX Tuv. and *i:yi* in XX Tuv. . The other area in which it has occurred is SW, and there perhaps only in Osm., Anat. and Rep. It is probably a recent development; that is certainly the impression given by the orthography. In Osm. written in Arabic script *y* < *g* was nearly always written with *ga:f* (*ka:f*), *ya*: being used only exceptionally. For example *ege* ‘a metal-worker’s file’ < *igeg* was so written in Osm., but in XIX, and perhaps somewhat earlier, was pronounced

¹ XIII Mong. did not tolerate a final unvoiced consonant and added a euphonic vowel to such words; other examples are *erke* < Turkish *erk*, *güçü* < Turkish *küç*, *kökö* < Turkish *kök* and some others mentioned below.

eye. Similarly VIII *beg* 'the head-man of a clan' was still so written in Osm., though pronounced *bey*. In Rep. orthography the sound is usually represented by *ğ*, e. g. *eğ*, pronounced *eye*, but *bey* is written as it is pronounced. In fact the sound change *-g* > *-y* is rare, and probably confined to monosyllables like *beg* > *bey*; normally *-g* simply disappears, as in *igeg* > *eye*. It should be added that the *-g* from which *-y* is evolved is sometimes itself a secondary sound; for example SW XX Rep. *ögle* (for *öyle*), 'noon', goes back through XIV ff. Anat. *ögle* to XI Xak. *ödleg*.

(5) Finally, it should be added, there are some cases of XX *y* which are 'irregular' in the sense that they have not evolved in any of the ways described above. For example, VIII *ev* 'dwelling place' (a neutral kind of word meaning tent, hut, house etc. as the local circumstances demand) has assumed some very odd and irregular forms in modern languages. In NE XIX these include *ü*.; *eb*, *ep*, *öp*, *üp*, *ög*, *üg*, *em*, *öm* and *üy*; in SE *öy*; in NC *ew*, *ü*.; *üy*; in SC *öy*, *üy* (in XIV *ew*); in NW *u*.; *üy* (pre-XX Kar. L., T. *iw*, *üw*, *yiw*, *yüw*, *yüw*); in SW consistently *ev* from the earliest period but with isolated occurrences of *öy* and *üy* in XX Anat... The course of evolution seems generally to have been *ev* > *üv* > *üw* > *ü*.: > *üy* with a euphonic *-y* attached.

I now turn the earlier history of initial XX *y*-. Of the four sounds from which XX *-y/-y* has evolved, only one, *y*, occurs as an initial in VIII Turkish; it might therefore reasonably be expected that XX *y*- would always go back to VIII *y*-. So far as pure Turkish words are concerned, I do not know of any case where this is not so; but there is a small group of Mong. loan-words (so far I have noticed about half a dozen and there are no doubt more) in which an initial *c*- in Mong., which goes right back to XIII and has been consistently retained in that language until the present day as *c*-, *dz*- and the like, has become *y*- in four or five NE XIX dialects (Alt., Bar., Küer. Leb. and part of Tel.).

The Mong. word *cida* 'a lance, spear' (synonymous with Turkish *süğü*.), which occurs as a loan-word in a wide range of Turkish languages from that of the XIV Uyğ.-Chinese Dictionary onwards, and even as far afield as Polish (*dzida*) and French (*gide*), is a good example. In SC XV ff. Çağ. and SW XIV Anat. right down to XX Anat. it has retained its original form *cida* (*sic*, not *çida*). The NE XIX forms are: Tuv. *cida*; Sag., Şor (and XX Xakas, Tuv.) *çida* (these languages devoice all initial *c*-s); Kç., Koib., part Tel. *t'ida*; Alt., Leb., part Tel. *yida*. It could of course be argued that the last mentioned languages and dialects alone have preserved a primitive Mong. sound which had already been altered in XIII Mong., but there are two excellent reasons for rejecting such a hypothesis. First the phonetic structure as a whole of these languages is so full of secondary sound changes that it is unreasonable to suppose that in this case they have been ultra-conservative; after all the sound change *c*- > *y*- has been noted in other languages. Secondly if the sound change *y*- > *c*- had already occurred in XIII Mong. the change should have been carried through consistently in all words as it has, for example, in XIX and XX Kır., Kzk. But in fact in XIII Mong. all three initials *c*-, *ç*- and *y*- exist, and they survive to the present day. In Prof. HAENISCH'S Wörterbuch zu Mangholun Niūça Tobça'an, Leipzig 1939 (quoted hereafter as HAENISCH) there are 12 pages of words with *c*-, 5 of words with *ç*- and 4 of words with *y*-. Given the smallness of the sample, this proportion is not significantly different from that in HALTOD'S Mongolian-English Practical Dictionary, 1949-53, that is 41-38-11, which seems to show that these initials have been pretty stable for the last 700 years.

Having thus traced the history of XX *y* backwards, I will now reverse the process and trace forwards the history of VIII *y* and the other VIII sounds from which XX *y* is derived.

(1) VIII *y*. While it is always dangerous to make confident statements that something never occurs, I can at any rate say that I have never found any case in which *-y/-y* have become some other sound, though intervocalic *-y-* and *-y-* at the end of a closed syllable are sometimes elided in the process of crasis. The history of *y-* is however quite different. Kaş. I 31 says that in Oğuz and Kıp. initial *y-* is either elided, as in *elkin* < *yelkin* 'guest' and *ılgı* < *yılgı* 'warm' or converted into *c-* as in *cinçü* < *yinçü*¹ 'pearl' and *coğdu*: < *yoğdu*:² 'the hair under a camel's neck'. It can safely be presumed that Kaş.'s *ci:m* in these words has the rare pronunciation *c*, instead of *ç*, to which he refers in I 8. The statement in I 31 is repeated in II 314 in the form that the "Türkmen" (sic) invariably have *c-* where the "Turks" (sic) have *y-*.

In modern times the position is rather more complicated; broadly speaking one of four things has happened; *y-* has been either retained, or converted into *c-*, or a secondary form of *c-*, or elided. A common word like *yat-* 'to lie down' (with various extended meanings) will serve as an example. *Yat-* remains unchanged in NE XIX Alt., Bar., Küer., Leb. and some Tel.; SE; NC XIX Tob. (only); SC and SW; in NW XIX Kaz. and XX Kk. both *yat-* and *cat-* occur; *yat-* has become *cat-* in NC XIX Kzk. and XX Kir.; *cat-* (*yat-*) has become *çat-* in NE XIX/XX Sag., Şor, Tuv. and *jat-* in NC XX Kzk.; finally in NE XIX Kç., Koib. and some Tel. the form is *t'at-*, which seems to be intermediate between *yat-* and *cat-*. To sum up *y-* is normally retained in a few NE, some NW and all SE, SC and SW languages, and altered to *c-*, or a further modification of that sound, in most NE and NC and some NW languages, but there are exceptions, perhaps due to inter-language borrowings, to this general rule.

¹ Kaş. is of course quite right in equating *cinçü* and *yinçü*, the latter a word as old as VIII Türkü, but there is good reason for supposing that this word is borrowed from the Chinese phrase *chên chu* (Giles 589-2549) (real) pearl. If so, *cinçü*, or rather *cincü*, was the original form, and *yinçü* a secondary one, and we have here a very early example of the change *c- > y-* just referred to. This has an important bearing on the problem discussed below whether VIII Turkish had an initial *c-* as well as *ç-*.

² This word presents rather a difficult problem. In this form it is unknown except in Kaş., but it is obviously identical with the synonymous (or practically so) word NE XX Tuv. *çogdor*, NC XX Kir. *coğdor* and NE XIX Şor *çoğdra* 'tassel', the last probably mistranslated and somewhat misspelt. It is not however possible to make the obvious textual amendment *ra*: for *wa:w* and make the word *coğdur* < *yoğdur*, since in III 30, in the list of words with long final vowels, the word is spelt quite clearly *yoğdu*: and in III 31 *yoğru*:/ *yoğru*: *y* is entered with the note that *r* is a secondary form of *d*. To add to the confusion *coğda*: appears with practically the same meaning in SC XIV Muh. and *çuda*, which must be a secondary form of that word in SC XIX Özb. (VÁMBÉRY'S Čagataische Sprachstudien) and NC XX Kir. VÁMBÉRY'S use of *c-* and *ç-* is so chaotic that no attention need be paid to his spelling, but Kir. carefully distinguishes between *c-* (< *y*) and *ç-* (< *ç*), and the Kir. *çuda* must have been borrowed from some language which devoices *c-*. Thus forms both with and without final *-r* are solidly attested (the former also by Proto-Bulgar, see below) and the only possible conclusion is that the original word had an unstable final *-r* which disappeared very early in some languages.

It will be noticed that the *y*- of *yat*- is nowhere elided, and in fact such elisions are rare and seem to occur almost exclusively before *-i*- and *-i*-. There is indeed some reason for supposing that Kaş. has got things the wrong way round and that in fact what happened was that a prosthetic *y*- was added before *i*-/*i*- in XI Xak. and a few other languages. The case histories of two very common words certainly point in this direction:

(a) *ıǵaş* 'a tree', so spelt in VIII Türkü, VIII/IX Man. A- and Ñ-dial., IX ff. Bud. and Civ. Uyğ.; NE *aǵaş* with nearly a dozen variants (*aǵas*, *aǵaş*, *aḡaş*, *ıyǵaş* etc.); NC XIX *aǵaş* (with variations in the final consonant); NW, SW consistently *aǵaş*; *yıǵaş* occurs only in XI Xak. and SC (but not consistently there) and *cıǵaş* in XX Kir.

(b) *ırak* 'distant', so spelt in VIII Türkü, IX ff. Bud. Uyğ.; NE *ırak*, *rak*; *ırak* also in SE, NC XX Kır., SC XIV Muh., NW XIV Kom.; *yırak* occurs only in XI Xak., NC XIX Tob., SC XV ff. Çağ. (*yırok* in XX Özb.), NW XIII-XV Kıp. and according to Radlov pre-XX Kar. L. T., but KOWALSKI has *ırax* for Kar.T.; *cırak* does not occur at all, which is most significant.

I cannot recall any case other than Kaş.'s *elkin* < *yelkin* of *y*- being elided before *-e*-, and if this is typical, the explanation may be that the *e* is a closed one, i.e. *élkin* < *yélkin*.

This does not exhaust the metamorphoses of *y*-; in some words in a few NE languages it becomes *n*- or *ñ*-. I have not completely analysed these words, but the languages concerned seem to be XIX Kç., Koib., Sag., and XX Xakas where it becomes *n*- and XIX Tub., where it becomes *ñ*-; *nama*-/*ñama*- < *yama*- 'to patch' is a good example. So far as I have examined them, all the words seem to contain a nasal later in the word and the phenomenon seems to be one of retroactive assimilation.

(2) VIII *d*. This sound occurs only in the medial and final positions. Kaş. I 32 says that "most of the Yağma, Tuxşı, Kıpçak, Yabaku, Kay, Çumul and Oğuz turn every *d* into *y* and never pronounce *d*; e.g. 'the *xalanc* tree' *kadıñ* > *kayıñ*; 'relation by marriage' *kadıñ* > *kayın*. Similarly some of the Kıpçak, Yemek, Suwar and Bulğar, that is the tribes stretching out to Rus and Rum pronounce the *d* in the language of the Çigil and other Turks as *z*; e.g. 'leg, foot' *adak* > *azak*, 'to be satiated' *tođ*- > *toz*-. To sum up, the Çigil *d* is *y* in Yağma, Tuxşı, Oğuz and some of the Arğu, going up towards China (*muşıdan nahwa*-'Si:n'), and *z* in Kıpçak and the others going down (*mustaffılan*) to Rum". The summary does not entirely agree with what goes before, and the text, with its double mention of Kıpçak, may be corrupt, but in effect Kaş. says that in XI *d* remained in the central area, but became *y* up China way and *z* down to the West. Muh. writing in XIV somewhere in the SC area says (p. 7 of MELIORANSKY's edition, p. 79 of RİFAT's) that "the Turkistanis" use *d* where "the Turks in our country" (elsewhere defined as "Iraq") use *t* or *y*, e.g. *idi* 'lord' > *iyi*; *edgü* 'good' > *eygü*; *adak* 'leg' and 'cup' > *ayak*. The MSS. are in a mess at this point, for example the crucial letter is called *da:l* not *da:l* and it is probable that the mention of *t* is a „Verschlimmbesserung“ by some scribe who thought that the subject under discussion was the devoicing of *d*, and not part of the author's text; the mistake that 'cup' (*ayak*) was once pronounced *adak* is also probably the copyist's and not the author's. The subsequent history of *d* is rather more complicated than could be foreseen by these authors and in one

respect rather different; there are very few traces of the sound change $\underline{d} > z$ in NW, though it is well established in some NE languages. The standard pattern, of which there are numerous examples, is that \underline{d} survives in SC until XIV, with occasional traces in XV, and is attested for one or two words in NW XIV "Bulğar" by Abu Hayyan's *Kita:bu'l-idra:k li-lisa:ni 'l-Atra:k* (edited by CAFEROĞLU, Istanbul 1931), but thereafter disappears, turning in a few NE languages into \underline{d} or z and everywhere else into y . Kaş. and Muh.'s own example *adək* 'leg, foot' is the best one to take. The recorded forms are:- VIII Türkü, VIII/IX Man. A-dial., Ñ-dial., Uyğ., IX ff. Bud. and Civ. Uyğ., XI & XIII Xak., XIV Uyğ.-Chinese Dictionary *adək*; XI Yağma etc. *ayak*; XI Kıpçak etc. *azak* (see above): NE XIX Alt., Bar., Küer., part Tel. *ayak*; Kç., Kız., Koib., Şor *azak* (XX Xakas *azax*); part Tel., Tuv. (and XX Tuv.) *adak*: SE pre-XX Türki *ayağ/ayak*: XIX Tar., XX Turfan etc. *ayak*: NC XIX & XX *ayak*: SC XIV Muh. *adək/ayak* (see above); Rbğ., Nahc *adək*; XV ff. Çağ. and later languages *ayak* (with *zak* 'foot' in PAVET DE COURTEILLE p. 109 which must be a muddled reminiscence of *adək*): NW XIII-XV Kıp. and XIV Kom. *ayak*; XIV Bulğar *adak*; pre-XX Kar. T. *ayax*; XIX Başkir (Radlov) *azak*; Krm. *ayak*; XX Başkir (KARIMOVA, Russko-Bashkirskiy Slovar', Moscow 1954) *ayak*: SW all periods *ayak*. It has been contended that the VIII Türkü word should be transcribed *adak* and that this should be regarded as the primaeval form, still surviving in NE XIX part Tel. and Tuv., but I hope that I have proved in my paper Turkish Ghost Words (J.R.A.S. 1955) that the Runic letter concerned represented both \underline{d} and \underline{d} , and I have pointed out in the discussion of the sound change Mong. $c > NE$ (part) y - that the NE languages are the exact reverse of ultra-conservative.

There are a few cases where words do not conform to this normal pattern; they may be classified as follows:

(a) $\underline{d} > -z$ at the end of a closed syllable is devoiced as in NE XIX Koib., Sag., XX Xakas *estig* 'rich' < IX ff. Bud. Uyğ. *edlig*. A more complicated case is *adğır* 'stallion'; VIII & VIII/IX Türkü, IX ff. Civ. Uyğ., XI Xak., XIV Uyğ.-Chinese Dictionary *adğır*; XI Oğuz *ayğır* (once in Kaş.): NE XIX Alt., Küer., Leb., Tel. *ayğır*; Kç., Koib., Sag., Şor., Tuv. *askır/aksır*, XX Xakas *asxır*: SE pre-XX Türki *ayğır/ağır*; XX Tfn. etc. *ayğır/ağır/arğır*: NC, SC, NW, SW all *ayğır*; as a loan word in XIII Mong. *acirxa* (compare (c) below).

(b) In a few words containing front vowels \underline{d} becomes $-g$ - (and sometimes later $-y$ -). A simple example is *eder* 'a saddle'; XI Xak. *eder*: NE *ezer/e:r*: SE *eger*: NC *eger/e:r*: SC *eger/eyer* (often spelt *iger/iyer*): NW *eyer*: SW XIV Anat. *eyer* (sometimes later spelt *eger*). Another word which shows this and other changes and produces an unusually complicated pattern, no doubt because when the \underline{d} disappeared it was so short, is *idi*: 'lord, master', which is also remarkable because it lost its \underline{d} - as early as IX ff. Bud. Uyğ. and survived as *idi* as late as XV ff. Çağ. . The forms which I have noted are VIII/IX Türkü *idi*: (the *idi*: in VIII Türkü is the other one, the adverb used with negative expressions to mean 'at all'); IX ff. Bud. Uyğ. *ige/iye/i*; XI Xak., *idi*:; NE *iye/e:/i*:; SE *ege/ige*; NC *ege/iye/e*:; SC XIV Muh. *idi* in Turkistan, *iyi* (misspelt *iti*) in Iraq; Rbğ., Nahc. *idi*; XV ff. Çağ. *idi/ige/eye/iye*; XIX Özb. *ege/eye*; NW XIV Bulğar *idi*, Kıp. *eye/iye*; XIX Kaz. *iye*; SW XIII/XIV and following Anat. *eye/iye/is* (or *is* or *ısı/isi?*), XX Anat. *eye/iye/ez*.

(c) In a few cases \underline{d} becomes $-c$ -. For example *adığ* 'a bear' for the most part follows the usual pattern, but becomes *ayığ/acu* in SE pre-XX Türki. Again *adruk* 'a grass with branching roots' (*Cynodon dactylon*, dhub-grass or the like), which

must be distinguished from *adruk* "different", is recorded in the forms XI Xak. *adruk*, Oğuz *ayrık*; SE pre-XX Türki *acruk*; SC XVff. Çağ. *acruğ*; SW Osm. and later *ayrık*. But the most remarkable in this class is *cacım* 'a kind of carpet or coverlet' recorded in some SW pre-XX Osm. authorities and found in XX Anat. as *cecim/cicim*. It must be a loan-word from Persian *ca:cim*, same meaning but that word is rightly described by HAIM (New Persian-English Dictionary, Teheran 1934) as a Turkish loan-word in Persian. It is in fact an altered form of XI Xak. *yadım* (Kaş. III 19 and other references), but from which language it was borrowed by Persian is obscure, since the sound change *y* > *c*- is characteristically Northern, while *-d* > *-c*- seems to be more Southern.

(d) There is one rather dubious case of *-d* > *-t/-d*, *öd* "time": VIII Türkü VIII-IX Man. N-dial. and Uyğ., IXff. Bud. and Civ. Uyğ., XI Xak., XIV Uyğ.-Chinese Dictionary *öd* (*ö:d*): NE XIX & XX Tel., Tuv. *öy*; SC XIV Muh. *öy*, Rbg. *öd* but SE XIX Tar. *öt* and SW XX Anat. (one reference only) *öd*. These two references are hardly a sufficient basis for a phonetic law, unless other instances can be found.

(e) Finally there is at least one case, and there may well be more, of "irregular" evolution. *Adut* 'a handful' is recorded at least once in IXff. Bud. Uyğ. and is listed in Kaş. I 50, with *awut* in I 83 as an alternative form. This is one of the very few cases of *w* in Kaş.; presumably the usual *-y*- was replaced by *-w*- under the influence of the *u* following. In nearly all modern languages the *-t* has become *-ç*, perhaps as a kind of reaction when the *-d*- disappeared, and there have been some very odd developments: - NE XIX Alt., Tel. *u:ş*; Kç., Koib., Sağ. *o:s*; Koib. *adıtya*; Küer. *auts*; Şor *o:ş*; Tuv. *adıyn* (RADLOV), *adış* (КАТАНОВ); XX Xakas *o:s*; Tuv. *adış*; SE pre-XX Türki *uwuç*; XX *ovuç*; XIX Tar. *o:ç*; NC XIX Kzk. *u:ş*; Tob. *u:ts*: NW *awuç/uwuç/uwuç* (but Kaz. *adış*): SW *avuç*.

(3) VIII *ñ*. This sound, except as a purely secondary form of *y*- in NE XIX Tub. (see above), occurs only in the medial and final positions and then very rarely. Two examples of *-ñ/-ñ* > *-y/-y* have been given above and Kaş.'s remarks on the subject quoted. The three examples which he gives of Xak. *y* > Arğū *n* are *koy/kon* 'sheep', *çiğay/çiğan* 'poor' and *kayū/kanu* 'what?' *Koñ* is relatively common in VIII and VIII/IX Türkü; *koy* is recorded for IXff. Uyğ., XI Xak., NE, SE, NC, SC (part) and NW (part), but SC XIV Muh. has both *koy* and *koyun* and NW XIII-XV Kıp. and SW (all languages) have only *koyun*; the word occurs as a loan-word in Mong. as *koni*, a very primitive form. In this case, therefore, besides *-ñ* > *-y*, the *-ñ* was broken up into *-ny* and the two parts metathesized. *Çiğan* is recorded for VIII and VIII/IX Türkü and *çiğay* for XI Xak. (Kutadğū Bilig as well as Kaş.) but the form in SC XIV Muh., XVff. Çağ. and NW XIV Kıp. is *çiğan*. *Kañu* is recorded only once in Runic script, oddly enough in VIII Uyğ., where it should theoretically be *kayū*; *kanyu*, so spelt, is recorded at least once for VIII/IX Man. N-dial.; *kayū* is recorded for VIII/IX Man. Uyğ., IXff. Bud. Uyğ., SC XIV Muh. and NW XIV Kıp. (as *kay*); the Locative Case, meaning 'where?' is recorded as *kayda* in NE, SE, NC, SC, and NW, and as *kanda* in IXff. Bud. Uyğ., XI Xak., SE, SC XIV Muh., XVff. Çağ., NW XIV Kıp. SW. Thus *-ñ/-ñ* > *-y/-y* and *-n/-n* seem equally common and *-ñ* > *-yun* also occurs.

(4) VIII velars and gutturals. The sound change velars and gutturals > *-y/-y* is only a small part of the history of this sound, which can only be written as part of the history of *k* (with front and back vowels) and *ğ/g*.

The conclusions of this part of my study can therefore be summarized as follows; the normal history is: -

VIII *y* > XX *y*-*c*-*ç*-*j*-*n*-*ñ*-*t*' (XIff. *y*- may sometimes be prosthetic).

VIII *-y*-*y* > XX *-y*-*y*.

VIII *-d*-*d* > XX *-c*-/(?not *-c*), *-d*-/(doubtfully *-d*, *-t*), *-s*-/(?not *-s*), *-y*-*y*, *-z*-/(?not *-z*).

VIII *-ñ*-*ñ* > XX *-n*-*n*, *-y*-*y*, and */yun*.

VIII velars and gutturals inter alia > XX *-y*-*y*.

XIII Mong. *c*- > XX *c*-*ç*-*t*'-*y*- in loan-words.

To the best of my belief this table is complete as far as it goes, that is to say I do not think that, apart from one or two irregularities, the sounds in the first column, except the velars and gutturals, ever turn into any secondary sounds other than those in the second column, or that XX *y*, except in a few exceptional words like *üy* < *e:v*, is a secondary sound of any VIII-XIII sounds except those in the first column; but obviously the other sounds in the second column are secondary forms of more sounds than those contained in the first column.

One interesting by-product of this research is further confirmation of the current belief referred to in the footnote on p. 39 above that there are no basic Turkish words with initial *c*-. The existence of medial and final *-c*-*c*, which is distinct *-ç*-*ç* and semantically significant, is of course now generally recognized, and also the fact that it is associated with a preceding long vowel, but owing to the complete impossibility of distinguishing between *c* and *ç*, if in fact both sounds existed in VIII-XI, in the Runic and Uyğur alphabets and in the Arabic script as it is used for transcribing the early texts, we cannot reconstruct the history of this sound and determine whether it is a *primaeva* Turkish sound which was devoiced in XI Xak. and some other languages, or whether it is a secondary phenomenon which emerged later, perhaps when differences between the length of vowels became blurred. In other words we know that in SW pre-XX Osm. *a:c* is "hungry" (and *od* 'fire') and *aç* 'open' (and *ot* 'grass'), we know that in XI Xak. the corresponding forms were *o:t* and *ot* and suspect that they were *a:ç* and *aç*, but we cannot tell whether in some pre-VIII form of Turkish the forms were *a:c*, *o:d*, *aç*, *ot* or *a:ç*, *o:t*, *aç*, *ot*.

Finally I come to the problem whether there is any means of carrying the history of *y* further back than VIII. It is at any rate suggestive that XX *-y*-*y* go back to VIII *-d*-*d* and *-ñ*-*ñ* as well as *-y*-*y*; logically XX *y*- too should in some words go back to *d*- and *ñ*- and not always to *y*-, even though by VIII these sounds, if they ever existed, had become *y*-. We must frankly face the fact that no help in solving the problem can be got from the languages so far studied. Those which maintain *y*- maintain it in all cases; those which convert it into *c*-, or before a nasal *ñ*- (both sounds secondary to *-y*-*y*), do so in all cases without exception, except those NW languages which use *y*- and *c*- indiscriminately for the same words. If evidence is to be found on this subject it must be found somewhere else. And in fact it can be found in two places, the far West and the earliest Turkish loan-words in Mong.

The evidence from the far West for the existence of a VI or VIII *d*- (> *y*-), though scanty, is conclusive. Menander Protector, a Byzantine author, records a VI Türkü word for 'a funeral feast', *doğia* (there is no doubt that at this period the Greek „delta“ was a voiced spirant dental and the „gamma“ a voiced guttural fricative); this is obviously VIII Türkü *yoğ* (or *yo:ğ*?), XI Xak. *yo:ğ* (Kaş. III 143) which has precisely this meaning. The other evidence comes directly or indirectly

from Proto-Bulgar and some of it is conveniently collected in PRITSÄK (1955) p. 43. In the first Bulgarian King-list (datable to late VII) the Snake-Year of the 12-year animal cycle is called *dilom* (certainly again a spirant *d*), which is obviously VIII/IX Türkü *yıla:n*. The Old Church Slavonic word *doxtor* 'a pillow' is probably a Proto-Bulgar loan-word and an earlier form with extended meaning of *yoğdu*: (see the note on p. 39). Similarly Magyar *dió* 'a nut' is believed to be derived from a Proto-Bulgar form of XI Xak. *yağa:k*, same meaning (Kaş. III 29). It is perhaps relevant that XIV Bulgar, as mentioned above, was a language in which -*d*- survived later than it did in most others.

The evidence regarding the earliest Turkish loan-words in Mong. is more difficult to disentangle, but easier with those taken from Turkish words with VIII *y*- than with most others. There are numerous loan-words with *y*- in the earliest Mong. translations of Buddhist texts, words like *yirtinçü* 'the universe', and it can safely be assumed that these were borrowed from IX ff. Bud. Uyğ., which was the language of the texts translated. But a careful study of HAENISCH's list of XIII words beginning with *y*- (rather over 40 in all) discloses only one word shared with Turkish, *yara* 'a wound', and that is probably a Mong. loan-word in Turkish and not vice versa. Its earliest occurrence in Turkish seems to be in NW XIV Kom., a language containing several other Mong. words, and it is otherwise unknown until it appears in SC XV ff. Çağ. Thus apparently before the 'Buddhist' period the Mongols were not in close touch with any Turks speaking Uyğ., Xak. and the other XI/XIII languages which retained *y*-.

On the other hand an examination of the words beginning with *c*- in HAENISCH's list, some 170 in number, discloses quite a number of words common to Mong. and the Turkish languages which have changed *y*- to *c*-. Some of them, words like *cida* 'a lance', *cerge* 'a row' and *coba*- 'to suffer' (with its derivatives) are clearly native to Mong. and loan-words in Turkish. But there are at least ten words which seem to me to be clearly native to Turkish and loan-words in Mong. These are: (1) *caka* 'a collar' from Turkish *yaka*: recorded in VIII Uyğ., XI Xak. and many later languages including NE, NC, NW *caka*/*çağa*; (2) *calbari*- 'to pray' from Turkish *yalvar*- in VIII/IX Türkü, IX ff. Bud. Uyğ., XI Xak., and many later languages including NC, NW *calbar*-; (3) *carxu* 'legal process', from Turkish *yarğu*, which I have not so far traced earlier than NW XIV Kom.; its Turkish origin is proved by the fact that, like the following word, it is derived from *yar*- 'to divide', a purely Turkish verb unknown in Mong.; (4) *carim* 'half' from Turkish *yarım* in IX ff. Bud. Uyğ., XI Xak. and many modern languages including NC, NW *carım*; (5) *carlix* 'royal decree' from Turkish *yarlığ* in VIII/IX Türkü, IX ff. Bud. Uyğ., XI Xak. and later languages including NC *carlık*; (6) *casa*- 'to arrange, ordain' from Turkish *yasa*- in VIII Türkü and many later languages including NW *casa*-; (7) *casax* 'ordinance', like *carxu* Turkish by etymology, though not recorded very early; (8) *cil*, used twice in dating formulae instead of the usual Mong. word (*h*)on, from Turkish *yıl* in VIII Türkü and nearly every Turkish language including NC, NW *cil*; (9) *ciruge*/*cürüge* 'heart' from Turkish *yürek* IX ff. Bud. & Civ. Uyğ., XI Xak. and numerous later languages including NC, NW *cürek*; for the final -*e* see the footnote to p. 37; (10) *cançi*- 'to strike' which is a case apart, since the immediate Turkish parallel, NC XX *canç*- goes back not to < *yanç*-, but to *sanç*- in VIII Türkü and many later languages. All these words must have been borrowed from some Turkish language, presumably in the NE group, which had already changed *y*- into *c*- by, or shortly before, XIII. This change is recorded in Kaş. as having already

taken place in XI in "Kıp. and Oğuz", a statement to be read in conjunction with his remark, quoted above, that *-d-* had become *-y-* in "the language of . . . the Oğuz . . . and others going up to China". If "Oğuz" is taken to have the same meaning in both cases, and if the point of departure is taken as Balasağun, the centre of Kaş.'s famous map, part of the area indicated is just about that where NE languages are now spoken.

Finally in my search for *d-* > *y-*, I examined HAENISCH's list of words beginning with *d-*, some 170 in all, to see if I could find any even earlier Turkish loan-words. As these borrowings must, by definition, have been made from a Turkish language more archaic than any we now know, precise phonetic equivalents with VIII ff. Turkish were not to be expected, but it seems to me that at any rate three words can reasonably be claimed as Turkish loanwords. They are: *dayi(n)* 'enemy', from Turkish *yağı*: in VIII Türkü & Uyğ., IX ff. Bud. Uyğ., XI Xak. etc. The sound change *y* < *ğ* occurs in other early loan-words, *ayıl* < *ağıl*, *takiya* < *takiğu* etc.; (2) *del* 'a horse's mane' from Turkish *yal/yel* in XI Xak. and all modern groups; (3) *dabçi* 'the cover' (of a pot etc.); there is no precise parallel to this word in Turkish, but it seems to be derived from Turkish (not Mong.) *yap-* 'to cover'; but there is also XIV Mong. *nabçi* 'a leaf' to be taken into account; this has been connected with Turkish *yapırğak*, same meaning, which is also derived from *yap-*; both etymologies cannot be right, since *yap-* cannot represent both < *dap-* and < *ñap-*. I should not at present like definitely to claim any other word in this list as Turkish, but I have my suspicions about *dura* 'heart', which might have been borrowed from an earlier form of *yürek* (see *cirüğe* above) and there may well be others. I have not attempted to comb the Mongolian dictionaries for other words not attested in XIII which should be taken into account, but Prof. PORRE has mentioned three to me: (1) *düyil/düyil-* 'to shave', contained in his list of words in the Mong. translation of the Muqaddimatu'l-adab (? XIV), which is clearly Turkish *yüli*:-, same meaning, XI Xak. and later languages; (2) *dulağan* 'warm' perhaps to be connected with Turkish *yılığ* (see above) if the *y-* in that word is not prosthetic; (3) *düri* 'form, shape, appearance', perhaps to be derived from Turkish *yüz*: 'face'; this is another problem of the *dabçi/nabçi* kind, for Mong. *nü'ur* 'face' is another candidate for the same etymology, but *yü:z* cannot represent both *dü:z* and *ñü:z*.

A hunt for proof of *ñ-* > *y-* through HAENISCH's list of words beginning with *n-* was frankly disappointing, particularly since about $4\frac{1}{2}\%$ of all the words in his list begin with *n-* as against a mere fraction of 1% in Turkish (excluding loan-words). The only really convincing case was *nodurxa* 'fist' from XI Xak. *yodruk* (Kaş. III 42) with the euphonic vowel attached (see footnote to p. 37). There may of course be others, but for the reasons mentioned above *nabçi* and *nü'ur* are doubtful cases, and it is difficult to connect both *nisun* "a tear" and *niray* 'fresh; young sprouts', or indeed either of them, with Turkish *ya:ş*, which has both meanings.

It is very likely that a hunt through the words in Magyar derived from Proto-Bulgar would be more productive. Prof. PORRE has pointed out to me that Magyar *nyár* "summer" is obviously derived from Turkish *ya:z* (< *ña:z*) through Proto-Bulgar *ña:r*.

To sum up this part of my study, there is clear proof, both from the far West and from Mong., that some words which in VIII began with *y-* in VI and earlier began with *d-*, and there is prima facie evidence, perhaps even proof, that other words which in VIII began with *y-* in VI and earlier began with *ñ-*. I should like to add that I have found nothing in the course of my researches to suggest that VIII *-y/-y*, *-d/-d*, and *-ñ/-ñ* represent different earlier sounds.