An Etymological

Dictionary of

Pre-Thirteenth-Century

Turkish

SIR GERARD CLAUSON

INDEX

PREFACE

1. In the first two chapters of *Turkish and Mongolian Studies*, Royal Asiatic Society Prize Publication Fund, vol. XX, London, 1962 (hereafter referred to as *Studies*), I gave a brief sketch of the history of the Turkish peoples and of their languages from the earliest period (the eighth century A.D.) at which they become directly known to us down to the medieval period, and attempted to identify and name the various dialects and languages which evolved during that period.

2. The broad thesis of this sketch was that a unitary Turkish language, which was not genetically connected with any other language known to us, and specifically not connected genetically with the Mongolian and Tungus languages, took shape, almost certainly in the steppe country to the west and north of the Great Wall of China, at some date which we cannot now determine, but certainly long before the start of the Christian era; that this unitary language split into two main branches, 'standard Turkish' and 'l/r Turkish', not later than, and perhaps before, the beginning of the Christian era; and that during the first millennium A.D. standard Turkish slowly broke up into two or three dialects, which soon became independent languages. No substantial early texts of any form of l/r Turkish have survived, but there are in Chinese historical works a few words of Tavğaç, the language of the Turkish tribe called by the Chinese T'o-pa, who were the ruling element in the Chinese Northern, or Yüan, Wei dynasty (A.D. 386–535), and there are some Turkish loan-words in Mongolian and Hungarian which were almost certainly borrowed from an l/r language, by the Mongols probably in the fifth or sixth centuries, and by the Hungarians probably in the ninth. The slight remains of Proto-Bulgar are in a similar language and so too are a few funerary inscriptions of the Volga Bulgars of the thirteenth and fourteenth centuries. The material is, however, very scanty and I have made only a few references herein to words in l/r Turkish, and those mainly where the only surviving l/r language, Chuvash (*Çuv.*), throws light on the phonetic structure of individual standard Turkish words. These citations are from N. I. Ashmarin, *Thesaurus linguae Tschuvaschorum*, 17 vols., Kazan–Cheboksary, 1928–50, cited as *Ash.* followed by the volume and page.

3. As I pointed out in *Studies*, p. 53, an essential basis for any scientific study of the comparative etymology of the Turkish languages is a list of all genuine Turkish words which are known to have existed before the various medieval Turkish languages were subjected to a mass invasion of loan-words from the languages of the foreign peoples (mainly Arabs, Iranians, and later Mongols) with whom the Turkish peoples then came into contact. The present work contains just such a list. I have endeavoured to include in it all the words known to have existed in the earliest known standard Turkish languages, that is Türkü,

Uyğur, including Uyğur-A, Old Kırğız, Xākāni, Oğuz, Kıpçak, and some minor
dialects, prior to the Mongolian invasion at the beginning of the thirteenth
century, excluding, except in a few special cases, words borrowed directly or
indirectly from the Indian, Iranian, and Semitic languages. All references to
'early Turkish' and 'the early period' relate to the Turkish languages as we know
them in the eighth to twelfth centuries inclusive.

4. One of the special characteristics of the Turkish languages is the manner in
which derived words are constructed from the basic words of the language by
a process of attaching suffixes to them. It is important, therefore, not merely to
compile a bare list of these early words and their meanings, but also to indicate
what kind of words they are (nouns, verbs, adverbs, etc.), verbs being distin-
guished from the rest by the attachment of a hyphen, e.g. at (noun) 'a horse',
at- (verb) 'to throw', and in addition, in the case of derived words, to analyse
their structure and indicate from what basic words they were derived, since this
often explains their semantic evolution. This is all the more important since,
even at the earliest date at which it becomes known to us, Turkish had already
had a long history, in the course of which some basic words had already become
obsolete, leaving the words derived from them, so to speak, 'in the air', with no
obvious origin; on this point see, for example, *Studies*, pp. 140 ff. In giving these
morphological explanations I have used the grammatical terminology customary
among English scholars of Turkish, eschewing the new terminologies which are
now taking shape in some academic circles.

5. My second purpose has been to establish the original meanings of all words
in the list and trace their subsequent semantic evolution by illustrating each
word by a quotation, or series of quotations, of passages in early texts in which
they occur. For this purpose I have examined all the surviving texts in the
languages enumerated above to which I have been able to obtain access. These
texts are enumerated in paras. 20-6 below.

6. The only words which have been included in the list without at least one such
quotation are a few which can first be traced in an actual Turkish text only in
the medieval period, that is during the thirteenth to sixteenth centuries inclusive,
but which demonstrably existed at an earlier period for such reasons as that
they exist both in standard Turkish and in Chuvash (in a distinctive form
peculiar to that language) so that they must already have existed before the split
between standard and l/r Turkish, or that words derived from them are attested
in the early period, or that they occur as loan-words in foreign languages in
forms which indicate that they must have been borrowed in, or before, the early
period. The texts which have survived from the early period are of course only
a fraction of those which once existed and there are no doubt other words which
existed in the early period but do not happen to occur in the texts which have
survived. Some of these words occur in medieval texts or still exist in modern
Turkish languages, but I have not attempted to identify and include them in the
list, since the evidence that they were really early words is lacking.

7. A surprisingly large number of early words are *hapax legomena* (Hap. leg.), that is, occur only once, and have not so far been discovered elsewhere; they have, however, been included in the list because they seem to be genuine Turkish words and may later be discovered elsewhere; but the fact that they are Hap. leg. makes it impossible to determine whether they are correctly transcribed. Properly speaking the term implies one single occurrence, but it has been stretched to include verbs listed in *Kaş.* with an enumeration of the Perf., Aor., and Infin. forms, since experience shows that if a word is mis-spelt in the Perf. it may also be mis-spelt in the Aor. and Infin.

8. When an early word has survived into the medieval period, I have also added a selection of quotations from the texts enumerated in paras. 27–52, and when it still survives in one or more modern languages I have usually recorded this fact also. There are two reasons for doing this: the first is that evidence of this kind is often required to fix the exact meaning and pronunciation of the word in the early period; the second is that, as I understand it, the main purpose of compiling a dictionary of this kind is not only to help students to read and understand the meaning of early texts, but also to provide a solid basis for the etymological dictionaries of medieval and modern Turkish languages which are now being compiled or will be compiled in future. I must, however, make it clear that my quotations from medieval texts and enumerations of modern languages in which a word survives do not profess to be as comprehensive as the quotations from early texts.

THE SYSTEM OF TRANSCRIPTION

9. The problems of transcribing early Turkish texts are discussed at length in *Studies*, Chap. 3 and elsewhere, and it is unnecessary to cover the whole ground again. There is, of course, no means of discovering exactly how Turkish was pronounced between 900 and 1200 years ago; the alphabets employed are all to a greater or less degree unscientific; the most that can be achieved is a broad system of transcription which will give an approximate idea of the pronunciation of the words concerned in the languages enumerated in para. 3. The transcription alphabet which I suggested in *Studies* and which is used here can best be described as the official alphabet of the Turkish Republic with one minor modification (regarding the use of ğ and g̣) and a few extra letters to represent sounds which either do not exist in Republican Turkish (Rep. Turkish) or, if they exist, are not represented by a distinctive letter. This alphabet was selected on the basis that the phonetic structure of early Turkish was probably much the same as that of Rep. Turkish, apart from certain sounds for which special letters have been provided. These sounds can be tabulated as follows:

10. *Vowels.* There are four back and five front vowels, all of which may be either short or long. The only additional letter is é for close e, a sound which

certainly existed in early Turkish and still survives at any rate in some dialects
of Rep. Turkish.

Back   {short **a i o u**
        {long **a: ı: o: u:**

Front   {short **e é i ö ü**
        {long **e: é: i: ö: ü:**

These are arranged, when circumstances demand it, in the following order:
a, a:, e, e:, é, é:, ı, ı:, i, i:, o, o:, u, u:, ö, ö:, ü, ü:; note that **u, u:** precede
ö, ö:. The round vowels represent a range of sounds which merge imper-
ceptibly into one another in the various languages.

In most, but not all, modern languages the vowels **o** and **ö** occur only in first
syllables of words, but there is good evidence, which I assembled in 'Three Notes
on Early Turkish', *Türk Dili Araştırmaları Yıllığı*, 1966, pp. 1 ff., that they were
also used in the second and later syllables of many words in early Turkish. This
evidence is, however, too fragmentary and imperfect, and in some cases too
inconsistent, to justify restoring such spellings at the beginning of all the entries
concerned. I have therefore entered all these words in the conventional spelling
with **u/ü** in the second and later syllables. Where there is good evidence in
regard to individual words that these vowels were originally **o/ö** I have entered
these spellings in brackets, e.g. **törü: (törö:)**, but I must emphasize that these
sounds occurred in far more words than are singled out for this treatment, for
example, in all words with the Suffix **-uk/-ük (-ok/-ök)**.

11. *Consonants.* The conclusion reached in *Studies* was that the sounds which
existed in early Turkish can be tabulated as follows, *v.* representing voiced and
u. unvoiced sounds:

	Plosive		Fricative		Nasal	Affricate		Sibilant		Semi-vowel
	v.	*u.*	*v.*	*u.*	*v.*	*v.*	*u.*	*v.*	*u.*	*v.*
Labial	**b**	**p**	**v**	**(f)**	**m**					**(w)**
Dental	**d**	**t**	**ḏ**		**n**			**z**	**s**	
Denti-palatal						**c**	**ç**			
Palatal					**ñ**			**(j)**	**ş**	**y**
Post-palatal	**ǵ**	**k**	**ʔ**		**ŋ**					
Velar	**ʔ**	**k**	**ğ**	**x**	**ŋ**					

Liquids (*v.*) **l r** (Aspirate (*u.*) **h**)

The post-palatal sounds occur only in association with front vowels and the
velar sounds only in association with back vowels, so only one letter, **k**, is
required to represent both unvoiced plosives. The unvoiced labial fricative **f**
probably did not exist in early Turkish as an independent sound, but seems to

have arisen as a secondary form of **v** in association with some unvoiced sounds. The voiced palatal sibilant probably did not exist as an independent sound in early Turkish, but is noted as a secondary sound in one or two words containing **ç**, e.g. **çöj-** probably for **çöz-**. It also occurred in a few Sogdian loan-words, e.g. **a:ju:n**, and a few other words, e.g. **üjme:**, which are probably, but not demonstrably, Sogdian loan-words. The labial semi-vowel **w** was not a native Turkish sound, but evolved as a secondary form of **v** in one or two languages like early Oğuz and also occurred in some loan-words. The voiced post-palatal **g** was normally a plosive, but may in certain positions have been a fricative; conversely the voiced velar **ğ** was normally a fricative but may in certain positions have been a plosive. The unvoiced velar fricative **x** probably did not exist as an independent sound in early Turkish, but seems to have arisen as a secondary form of **k** in association with some sounds, and also occurred in a few loan-words, some of them very old like **xağan** and **xan**. Kāşğarī says that the simple aspirate **h** was not a native Turkish sound but occurred in some 'impure' dialects and in one or two words like **ühi:** in which it seems to be a secondary form of **g**.

The consonants are arranged, when the circumstances demand it, in the following order: **b, p, v, (f), (w); c, ç, j; d, ḍ, t; ğ,** (velar) **k, x; g,** (post-palatal) **k, (h); l; m; n, ŋ, ñ; r; s; ş; y; z.** Note that the letters are arranged in classes, and not in the normal conventional order, but that the first letters of the classes are arranged in the conventional order.

12. There are some particular difficulties in transcribing some early languages, of which the following are the most important:

(1) In Türkü there was a tendency towards dissimilation, so that when a suffix beginning alternatively with a voiced or unvoiced consonant like the Perf. suffix **-dı:/-tı:/-di:/-ti:** was attached to a verb ending in a consonant, the unvoiced form was attached to a voiced consonant and vice versa, e.g. **erti:, tapdı:.** In Xak. the position tended to be exactly the opposite, at any rate to the extent that suffixes with voiced initials were attached to verbs ending with voiced consonants, e.g. **erdi:,** although cases of suffixes with voiced initials being attached to verbs ending in unvoiced consonants are very common, at any rate in our manuscript of *Kaş.* Chronologically speaking Uyğ. lay squarely between Türkü and Xak. but nearly all the alphabets in which it is written are so ambiguous that it is quite uncertain whether it was a dissimilating language like the former or an assimilating language like the latter, or whether the change took place at some date between the eighth and eleventh centuries. I cannot therefore claim to have been in the least consistent in transcribing Uyğ. texts, usually following the scriptions in the published texts.

(2) The Runic alphabet (see *Studies,* pp. 51 ff.) in which nearly all the Türkü texts have survived, the Manichaean Syriac alphabet, and the Brāhmī alphabet had different letters for **b** and **p** and for **g** and post-palatal **k**, the other early alphabets did not, and the Brāhmī spellings are very erratic; there is indeed some doubt regarding the phonetic value of some of the letters of this alphabet

in its Central Asiatic form. It is therefore almost impossible to decide how some words should be transcribed. In some cases assistance can be got from the modern languages, but some of these have undergone quite considerable phonetic changes.

(3) Only one of the alphabets concerned, Brāhmī, had separate letters for **o, u, ö** and **ü**, and the spellings of the texts in this alphabet are often inconsistent. In this case, however, the correct form can often be established if the word was a loan-word in Mongolian, or survives in modern languages; but the spellings in modern languages are often inconsistent, particularly as between languages in the Oğuz group and the rest.

13. The alphabet set out above has also been used for transcribing (or perhaps it would be more accurate to say transliterating, since the system adopted is purely mechanical) quotations from Arabic and Persian texts, with the addition of the following letters: **t, d, ṭ, ḥ, ṣ,** and **ẓ,** *hamza* ', *'ayn* ' and **q** for the unvoiced velar plosive, which must be distinguished from the unvoiced post-palatal plosive in these languages. In these languages, too, vowel length is indicated in the usual manner by a superscribed line, e.g. **ā.** The same letters are also used to transcribe Mongolian, but Sanskrit is transcribed in the conventional way.

THE ARRANGEMENT OF WORDS IN THE DICTIONARY

14. As pointed out above, the Turkish texts quoted in this book are written in a variety of alphabets, all more or less ambiguous, and it is often impossible to determine the correct transcription of a number of words; moreover, some words were pronounced slightly differently in different languages. It would, therefore, not be sensible to arrange the words in the strict alphabetical order to which we are accustomed in the dictionaries of European languages, since this would involve a great many double or multiple entries and greatly add to the difficulty of finding individual words. This problem and the means of solving it are discussed in detail in *Studies*, Chap. 4. The arrangement which emerged from this discussion can be summarized as follows:

(1) The words are broken down into fourteen groups, containing respectively words beginning with the following sounds:

(*a*)	vowels	(*h*)	**m**
(*b*)	labial plosives (**b, (p)**)	(*i*)	**n**
(*c*)	denti-palatal affricate (**ç**)	(*j*)	**r**
(*d*)	dental plosives (**t, (d)**)	(*k*)	**s**
(*e*)	velar plosives (**k, (x)**)	(*l*)	**ş**
(*f*)	post-palatal plosives (**k, (g)**)	(*m*)	**y**
(*g*)	**l**	(*n*)	**z**

(2) Each of these fourteen groups is divided into sub-groups; in each case the first sub-group contains monosyllables ending in an open vowel; the order of the

remainder is determined in the case of the first group (words beginning with vowels) by the first consonant, and, if there is one, the second consonant or failing that the final vowel, if any, and in the case of other groups by the second consonant, and the third consonant, if any, or failing that the final vowel, if any. For purposes of classification these are the only significant sounds; vowels other than initial and final vowels are disregarded except for purposes of arranging the order of the words in a sub-group.

15. Each sub-group is identified by from one to three code letters representing the significant sounds enumerated above, the code letters being as follows:

A, or if one of the significant sounds is post-palatal
ġ or k, E, representing vowels.
B representing labials, b, p, v, (f), (w) but not m.
C representing denti-palatal affricates c, ç, (j).
D representing dentals d, ḍ, t but not n.
Ğ representing velars ğ, k, x but not ŋ.
G representing post-palatals ġ, k, (h) but not ŋ.
L representing l.
M representing m.
N representing n, ŋ, ñ.
R representing r.
S representing s.
Ş representing ş.
Y representing y.
Z representing z.

Each sub-group is divided into the following parts:
(1) monosyllables, if any, coded as Mon.
(2) monosyllabic verbs, if any, coded as Mon. V.
(3) disyllables, coded as Dis.
(4) disyllabic verbs, coded as Dis. V.
(5) longer words, coded as Tris.
(6) longer verbs, coded as Tris. V.

16. In order to locate a word in the dictionary:

(1) declensional and conjugational suffixes must be removed;
(2) the part in which it appears in the dictionary must be determined by converting the significant sounds in it into the appropriate code letters, account being taken of the fact whether it is either a verb or not a verb, e.g.:

u:	Mon. A
u:-	Mon. V. A-
ak, oğ, u:ğ, ok	Mon. AĞ
eg-, ek-, ög-, ük-	Mon. V. EG-
bu:	Mon. BA

çars	Mon. CRS
tüpi:, töpü:	Dis. DBA
tapın-, tevin-	Dis. V. DBN-
tügü:nük	Tris. DGN
tıltağlığ	Tris. DLD
yarlıka:-	Tris. V. YRL.-

17. Within each part of a sub-group the words are arranged in the following order, which is designed to make it as easy as possible to locate a word written in any of the various scripts:

(1) in Mon. and Mon. V. in the alphabetical order of the vowel, and when two or more words have the same vowel (o and u, and ö and ü, respectively being regarded as identical since they are not distinguished in these scripts) in the order of the final consonants; thus in Mon. V. AD- the order is aḏ-, at-, ét-, ı:ḏ-, it-, u:ḏ-, ut-, üḏ-, öt-, üt-.

(2) in Dis. and Dis. V.,

(a) all words ending in a closed syllable precede those ending in an open vowel, thus in Dis. V. ADL- ütül- precedes atla:-;

(b) if two words contain, in addition to the letters represented in the heading, further consonants belonging to different classes, these determine their order; thus in Dis. V. SCD- sıçtur- precedes saçtaş-;

(c) in other cases, subject to (a) above, the order is determined by the first vowel; thus under Dis. V. SCD- the order is saçıt-, sıçıt-, suçıt-, süçıt-, and under Dis. V. ADL- atla:-, edle:-, or, if the vowels are identical, by the order of the consonants following, thus under Dis. V. ADL- oḏul, uḏul-, otul-.

(3) in Tris. and Tris. V. the same principles are observed, *mutatis mutandis.*

18. Most words in the dictionary are preceded by code letters in capitals, only basic words of which the pronunciation is reasonably certain not being preceded by one. These letters have the following meaning:

(1) C compound; used when a word is composed of two shorter words, e.g. C içkur 'belt', composed of iç and kur. Such words are rare.

(2) D Derived; placed before every word other than a compound which is not a basic word, e.g. D evdil-, the Passive form of evdi:-.

(3) F Foreign: placed before words which are known to be loan-words, e.g. F a:ju:n borrowed from Sogdian. Words which are believed to be loan-words are preceded by ?F if definite proof is lacking.

(4) PU Pronunciation uncertain; see (6).

(5) S Secondary; placed before words which are secondary forms of other words in the list; such words are followed by a reference to the words of which they are the secondary forms.

(6) VU Vocalization uncertain. This prefix is used when the vocalization of a word is uncertain but there is no doubt about its consonantal structure, e.g.

VU **eteç** indicates that this is the most probable pronunciation of the word but that it might be **ataç, atıç, atuç,** etc. Words are preceded by PU when there is some doubt about the pronunciation of one or more consonants in them whether or not the vocalization is certain.

(7) When a word is known to have existed at an earlier date but was already obsolete by the eighth century so that it is known only from its derivatives the basic word is entered with a preceding asterisk and followed by cross-references to some of the words derived from it, e.g. 1 *a̠d-, see a̠dın, a̠dır-.

(8) E Erroneous; unfortunately a few completely non-existent words have found their way into sóme medieval and modern dictionaries and word-lists owing to misreadings of ancient texts. A small selection of these is included in the list, preceded by E and followed by an explanation, in order to indicate that such words really did not exist and have not merely been overlooked, e.g. E **amuç** which arises from an erroneous scription of the Arabic word *aṣaḥḥ* in the manuscript of *Kaş.*

AUTHORITIES QUOTED, WITH SHORT TITLES

I. EARLY TEXTS

19. All quotations in the paragraphs about individual words are preceded by the name of the language in which they were written and the date of composition expressed in centuries denoted by roman numerals: VIII = eighth century A.D. and so on. When two or more quotations are taken from the same language they are grouped under a single head, possibly with different chronological indications.

A. *Türkü*

20. Two periods are distinguished:

(1) VIII. Several major inscriptions in the Runic alphabet fall under this heading; there is a brief note on the best editions of them in *Studies*, p. 68, but my own quotations are derived, wherever possible, from a personal examination of the photographs and squeezes of the inscriptions. As a result in one or two cases I have, I think, found better readings of some difficult passages than those in any of the editions; examples will be found under **olğurt-, berüki, sın-.** In addition to the major inscriptions listed separately below there are some shorter inscriptions, but as they are undated and may be later than VIII they are included under the next heading. The major inscriptions are the following:

(*a*) the funerary inscription of Toñukuk (this is the customary transcription of the name but it is VU) composed probably in the second, but possibly the third decade of VIII; cited as *T* followed by the number of the line;

(*b*) the funerary inscription of Kül Tégin, composed in A.D. 732, cited as *I* followed by the side (E, W, N, S) of the stone and the line on the side;

(*c*) the funerary inscription of Bilge: Xağan, composed in A.D. 735, cited as *II* similarly followed;

(d) the funerary inscription of the Küli Çors at Ixe-Xuşotu in central Mongolia, roughly contemporary with the preceding, cited as *Ix.* followed by the line;

(e) the relatively brief inscription at Ongin, cited as *Ongin* followed by the line.

(2) VIII ff. Under this heading are included texts which may have been composed in VIII, but were possibly, in some cases probably, composed in IX, or perhaps in one or two cases even later. These fall under the following heads:

(a) a few minor inscriptions from Mongolia, of which the most convenient editions are those in the second volume of H. N. Orkun, *Eski Türk Yazıtları*, 3 volumes and index, Istanbul, 1936–41, (cited as *ETY* followed by the volume in roman and the page in arabic figures). These are occasionally cited under the names given in that volume followed by a reference to *ETY II*;

(b) those of the so-called Yeniseian (Yen.) inscriptions which were found in Khakassia and were probably erected by minor Western Türkü chieftains who escaped to Khakassia after the fall of the Türkü empire in the middle of VIII, see *Studies*, p. 69. The most convenient edition is that in S. Ye. Malov, *Yeniseiskaya pis'mennost' tyurkov*, Moscow, 1952; they are cited as *Mal.* followed by the number of the inscription and the line, but my readings sometimes differ from Malov's;

(c) a few inscriptions found at Ayrtam Oy near the town of Talas on the river of the same name and probably datable to IX or X (see *Studies*, pp. 72 and 256). The texts are almost illiterate and in a deplorable state, and I have quoted them only occasionally, citing them as *Talas* followed by the number of the inscription and the line;

(d) paper manuscripts in the Runic alphabet. The only substantial one is a book of divination called *Irk Bitig*, which I cite as *Irk B* followed by the number of the paragraph and not, as has usually been done, by the page, since the page numbers used do not correspond to those of the actual manuscript. The most convenient edition of the rest, which are rather miscellaneous in character, is in *ETY II*, but as Orkun's system of references is rather chaotic I have substituted my own, followed by a reference to the page in *ETY II*. It is commonly believed that these documents can be dated to about IX;

(e) Manichaean (Man.) texts in the Manichaean Syriac or Uyğur alphabets and what Prof. A. von Gabain in her *Alttürkische Grammatik*, Leipzig, 1941, p. 5 (cited as *v. G. ATG*) called 'the N-dialect', but which seems in fact, see *Studies*, p. 118, to be Türkü written in a non-Runic alphabet. The longest and most important of these texts is the 'Confession of Sins' known as the Chuastuanift (more correctly Xwāstwanēft). The best manuscript, almost complete, is in the British Museum, and there are other fragments, some of the beginning of the text missing in that manuscript, in Germany. The most convenient edition, with a facsimile of the B.M. manuscript, is still that by A. von Le Coq in JRAS 1911, pp. 279 ff., although the translation is not entirely accurate. I cite it as *Chuas.* followed by the line of the B.M. manuscript or, in the case of the first part, as *Chuas. I* followed by the line in von Le Coq's numeration. The other

Manichaean Türkü texts which I quote have all been published in the *Abhand-lungen* or *Sitzungsberichte* of the Academy of Sciences in Berlin (originally the Königliche preussische, later Preussische, and later still Deutsche, Akademie der Wissenschaften). As these publications also contain Manichaean texts in Uyğur and Uyğur-A it is more convenient to list all the publications of Mani-chaean texts here to avoid additional references in subsequent paragraphs. The main bulk of them is in A. von Le Coq's *Türkische Manichaica aus Chotscho* I (AKPAW, 1912); II (APAW, 1919); III (APAW, 1922), cited as *M I, II,* and *III* followed by the page and line, the last followed by (i), (ii), etc. if the same line number appears more than once on the same page. Others were published in a series called *Türkische Turfantexte* (cited as *TT*), edited at one time or another by W. Bang, A. von Gabain, and G. R. Rachmati (later R. R. Arat), which also contains Uyğur Buddhist and Civil texts. Those parts which contain Manichaean texts (mostly not in Manichaean Türkü) are *TT II* (SPAW, 1929) cited by page and line, and *TT III* (SPAW, 1930) and *TT IX* (ADAW, 1958) both in Uyğur and cited by line. Other short Manichaean texts are in A. von Le Coq, *Ein manichäisch-uigurisches Fragment aus Idiqut-Schachri* (SKPAW, 1908) cited as *Man.-uig. Frag.* followed by page and line, and W. Bang and A. von Gabain, *Ein uigurisches Fragment über den manichäischen Windgott* (Ungarische Jahrbücher VIII, pp. 247 ff.), cited as *Wind.* followed by the line.

B. *Uyğur and Uyğur-A* (*Uyğ. and Uyğ.-A*)

21. The reasons for supposing that two separate, but closely related, Uyğ. dialects existed are stated in *Studies*, p. 42. There are obvious traces of Uyğ.-A in some Buddhist texts, e.g. those in *TT VI* and *TM IV* (although it is stated in the introduction to the latter that the Uyğ.-A forms had been 'corrected' in the transcription so that they are no longer apparent); but such texts do not seem to be entirely homogeneous, so that it is only in the Manicheaean texts that it seems possible more or less to isolate the two dialects. Three periods can be distinguished:

(1) VIII. The earliest and most substantial Uyğ. inscription in the Runic alphabet is that situated at Şine-usu in Outer Mongolia, which is the funerary monument of an Uyğur Xağan who ruled from A.D. 746 to 759. It is cited as *Şu.* followed by the side and line on the side.

(2) IX. The second such inscription, only fragments of the Uyğ. text of which still survive, is the funerary memorial at Kara Balgasun of an Uyğur Xağan who probably ruled from A.D. 808 to 821. It is cited as *III* followed by a letter and line and a reference to *ETY II*. The third is the short funerary inscription at Suci in Outer Mongolia, probably datable to the first quarter of IX, and cited as *Suci* followed by the line.

(3) VIII ff., that is, in this case, texts the earliest of which may go back as far as VIII, although the actual manuscripts are probably later, while the latest, apart from the XVIII manuscript of the *Suvarṇaprabhāsa Sūtra*, the composition of which was much earlier, is probably a Civil document dated firmly to the second

lines 1 to 324, the second lines 1763 to 2160, the rest being unpublished); and the Turkish translation of the *Suvarṇaprabhāsasūtra*, published by V. V. Radlov and S. Ye. Malov, St. Petersburg, 1913 ff., cited as *Suv.* followed by the page and line.

(*d*) Civil (Civ.) texts, that is not religious. These fall into four classes:

(i) Medical texts, probably mere translations of texts in other languages. The main collection is in G. R. Rachmati, *Zur Heilkunde der Uiguren* I (SPAW, 1930) cited as *H I* followed by the line, and II (SPAW, 1932) cited as *H II* followed by the page and line. There are also one or two such texts in *TT VII* and *VIII*.

(ii) Astronomical, astrological, magic, and omen texts, published in *TT I* (SPAW, 1929), cited by the line, *TT VII* and *VIII* and *USp.* Texts of the first two kinds are mainly translated from the Chinese, generally with an Indian background, and are mostly late in the period; the line between unorthodox Buddhist and purely magical texts is indefinite and I may have put some texts the wrong side of it.

(iii) Commercial and legal documents and (iv) miscellaneous texts. The main collection of these is in *USp.*; there is one in *TT VII*, No. 42, a very late text apparently transcribed from an original in the Arabic alphabet, and there are one or two published elsewhere. When quoting the last I have indicated where they may be found.

A good many of the Uyğ. texts listed above have been completely or partially indexed; but no index is available for others, including *Suv.* which is much the longest, and I cannot guarantee that I have listed every word which occurs in these unindexed texts.

C. Old Kırğız (O. Kır.)

22. It is commonly believed that those of the Yeniseian inscriptions which were found not in Khakassia but in Tuva were set up by Kırğız chieftains in that area and are in the Old Kırğız language. It has been suggested that they are the oldest monuments in the Runic alphabet and may date back as far as VI, but L. R. Kyzlasov has recently proved by archaeological methods (see *Studies*, p. 70) that most of them date from IX and X and that some may be even later. I have classified them as IX ff. The most convenient edition is that of Malov mentioned in para. 20 (2) (*b*), but A. M. Shcherbak has published revised texts of some of them, and some photographs and squeezes are available. I have therefore not always followed Malov's readings.

D. Xākānī (Xak.)

23. Xak. is a language closely related both to Türkü and to Uyğ., but sufficiently distinct from both to be regarded as a separate language. It was certainly not directly descended from the latter, indeed it existed side by side with Uyğ. for two or three centuries, and was perhaps not quite directly descended from the former. It is first identified in XI, and there are two Xak. XI texts of major

quarter of xiv. The actual date of composition of the remainder is quite un-
certain; some of the legal and commercial documents are certainly not older
than xiii and contain Mongolian, Arabic, and Persian loan-words, but I have
taken them into account with suitable precautions, since it is quite impossible
to divide these documents neatly into two categories, those anterior and those
subsequent to A.D. 1200. These texts can be divided into the following classes:

(*a*) Christian texts (*Chr.*). These are likely to be quite early in the period.
Only two seem to have survived, the well-known apocryphon about the Magi
published in F. W. K. Müller, *Uigurica* (AKPAW, 1908) pp. 5 ff., cited as *U I*
followed by the page and line, and a fragment published in *M III* cited by page
and line.

(*b*) Manichaean (*Man.*) texts in Uyğ. and Uyğ.-A published in the various
works listed in para. 20 (2) (*e*) above. All are fragmentary and vary greatly in
length, though none are very long; in the case of the longer texts it is usually
possible to determine whether they are in Man. Türkü, Uyğ. or Uyğ.-A; when
this is impossible they are cited as *Uyğ*. The earliest of these texts probably go
back to viii and the latest are unlikely to be much later than ix.

(*c*) Buddhist (*Bud.*) texts. These have been published in substantial quantities;
the following is a list of those which I have quoted. There are four volumes of
Uigurica, the first three edited by F. W. K. Müller, the last by A. von Gabain.
U I has already been mentioned under (*a*); *U II* (AKPAW, 1911), *U III*
(APAW, 1922), and *U IV* (APAW, 1931) are cited by page and line. The follow-
ing volumes of the *TT* series mentioned in para. 20 (2) (*e*) contain Buddhist texts:
TT IV (SPAW, 1930); *TT V* (SPAW, 1931); *TT VI* (SPAW, 1934); *TT VII*
(APAW, 1937); *TT VIII* (ADAW, 1954, the volume containing the texts in
the Brāhmī alphabet), and *TT X* (ADAW, 1959). Of these *VI* and *X* are cited
by the line, *IV* and *V* by page and line, and *VII* and *VIII* by document and
line. There is a useful and complete *Analytischer Index zu den fünf ersten Stücken
der TT* (SPAW, 1931) and *VI* contains a partial and *VII*, *VIII*, and *X* a com-
plete index of words. Another AKPAW volume containing Buddhist texts is
F. W. K. Müller, *Zwei Pfahlinschriften aus dem Turfanfunden*, 1915, cited as
Pfahl. followed by the page and line. W. Radloff, *Uigurische Sprachdenkmäler*,
Leningrad, 1928, cited as *USp.* followed by document number and line, con-
tains *inter alia* a number of Buddhist texts. Other publications which I have
quoted are the following: Paul Pelliot, *La Version ouigoure de l'histoire des
princes Kalyānamkara et Pāpamkara*, T'oung Pao XV. 2, 1914, cited as *PP*
followed by the page and line; W. Bang and G. R. Rachmati, *Türlü Cehennemler
üzerine Uygurca Parçalar*, Türkiyat Mecmuasi IV, Istanbul, 1934, cited as
TM IV followed by the line; Şinasi Tekin, *Kuanşi Im Pusar*, Erzurum, 1960
(superseding Radloff's partial edition, St. Petersburg, 1911) cited as *Kuan.*
followed by the line; W. Radloff, *Tišastvustik, ein in türkischer Sprache bearbeitetes
buddhistisches Sutra*, St. Petersburg, 1910, cited as *Tiş.*, followed by the manu-
script page and line; A. von Gabain, *Die uigurische Übersetzung der Biographie
Hüen-tsangs* (SPAW, 1935) and *Briefe der uigurischen Hüen-tsang-Biographie*
(SPAW, 1938), both cited as *Hüen-ts.* followed by the line (the first containing

importance. The first is the earliest, and by far the most important, Turkish–Arabic dictionary, the *Dīwān Luǧāti'l-Turk* of Maḥmūd al-Kāšǧari, dating from the third quarter of XI and cited as *Kaš*. The quotations in the form *Kaš.* followed by a roman number and an arabic number and sometimes a second arabic number are from the Turkish translation by B. Atalay, *Divanü Lugat-it-türk Tercumesi*, 3 volumes and index, Ankara, 1940–3, since this edition is reasonably accessible, and the index is excellent and comprehensive. The principle which I have followed is to cite the volume and page for main entries and the line as well for quotations from other parts of the book. I have, I hope, included every word which it contains. Atalay's translation is not wholly satisfactory and I have in every case consulted the facsimile (the printed text contains many minor inaccuracies) and normally quote the original Arabic in my citations, so that the accuracy of the translation can be checked. In a number of cases I have altered Atalay's readings and spellings. The second major text is the *Kutadǧu: Bilig* of Yūsuf Xaṣṣ Ḥācib of Balasagun, the earliest surviving purely literary text in any Turkish language, also dating from the third quarter of XI. My quotations are taken from R. R. Arat's critical edition, Istanbul, 1947, cited as *KB* followed by the line in that edition. In cases of doubt I have also consulted the facsimiles of the three manuscripts; in some cases my spellings are slightly different from Arat's. As his index has not yet been published, I cannot guarantee that I have listed every word which occurs in this text, but I have searched it fairly thoroughly. I have classified all the words from these two sources as XI since this date is certain. The Verse Preface of *KB* is not part of the original text and its language is obviously a little later; quotations from it have therefore been entered under the heading of XII (?) *KBVP* followed by the line. The Prose Preface is later still and perhaps not really Xak. but a few quotations from it have been entered as XIII(?) *KBPP* followed by the line.

E. *Oğuz*

24. Kāšǧari lists a certain number of words in a language which he calls 'the language of *al-Ǧuzziya*' and says that owing to long contact between the people who spoke it and Persian-speakers it had acquired a number of Persian loanwords, often in a rather corrupt form (see **ören** and **turma:**). It was no doubt the language spoken by the Oğuz tribes, who were then living in the neighbourhood of the Aral Sea and had to some extent lost contact with the other Turkish tribes. They were descendants of the Oğuz (Tokkuz Oğuz etc.) who had played quite an important part in history during VII and VIII and probably even earlier. There is no reason to suppose that during that period they spoke a language markedly different from that of the Türkü; but although even by XI the Oğuz language had undergone some phonetic decay and accepted a number of loanwords, it retained, and its modern descendants still retain, some characteristics more primitive even than those of Türkü, e.g. initial **d-** in certain words (see para. 14 (1)). Only a few specifically Oğuz words are listed in *Kaš.* but they are the earliest surviving specimens of the language.

F. *Kıpçak (Kıp.)*

25. Kāşğarī also cites a number of words in a language which he calls 'the language of Qifcāq', a language which he describes as similar to Oğuz. Several words are mentioned as being common to Kıp. and Oğuz but not known in Xak. In XI the Kıpçak were west of the Oğuz in southern Russia and *Kaş.* is the earliest authority for their language.

G. *Minor eleventh-century dialects*

26. Kāşğarī describes a few words as belonging to other contemporary dialects with which he was acquainted; those mentioned most often are the dialects of Arğu:, Çigil, and Gancak (the last probably not really Turkish, see *Studies*, p. 132 and the remarks on **kendük** below). I have included these words with Kāşğarī's descriptions of them.

II. MEDIEVAL TEXTS

27. The Uyğ. language lingered on for some centuries in Chinese Turkistan (Sinkiang); indeed the latest surviving Bud. Uyğ. manuscript, that of the *Suvarṇaprabhāsasūtra* already mentioned, was copied in XVIII, but the only accessible authority for Uyğ. in this latest state is the Chinese–Uyğ. dictionary (*Chin.–Uyğ. Dict.*) prepared by a committee of Chinese scholars towards the end of XIV. Some words from this work are included in Radloff's *Wörterbuch* and a complete list from another edition was recently published by Prof. L. Ligeti in *Un Vocabulaire sino-ouigoure des Ming, le Kao-tch'ang-kouan Yi-chou du Bureau des Traducteurs*, Dissertationes Sodalium Instituti Asiae Interioris II, Budapest, 1966. I have included references to this work as *Ligeti*, followed by the page.

28. The difficulty about classifying most of the remaining medieval texts, which were, with only one or two exceptions, written by or for Moslem Turks, is that in XI the only literary language in which texts have survived, and probably the only literary language then in use, was Xak., but that towards the end of that century or very soon afterwards literary languages emerged which were written in different areas where the spoken languages were rather different from one another, and so very soon began to pursue different courses. There are interesting discussions of the problem of classifying the texts concerned in *Philologiae Turcicae Fundamenta* I, Wiesbaden, 1959 (cited as *Fundamenta*) and more particularly in J. Eckmann's article 'Das Chwarezmtürkische' in that volume, and also in the introduction to A. K. Borovkov, *Leksika sredneaziatskogo Tefsira XIII—XV vv.*, Moscow, 1963 (cited as *Tef.*).

29. What can perhaps be called the direct line of descent, or southern stream, evolved in the area called in the early medieval period Khorasan and the southern part of Mā warā'u'l-nahr (Transoxiana). By XV the literary language used in that area was what is commonly, if inaccurately, called Çağatay (*Çağ.*), but at any

rate three, and probably four, texts have survived which were written between
XI and XV (both excluded) in languages intermediate between Xak. and Çağ.
The spoken language in this area during this period seems to have been fairly
homogeneous with the literary language, except that there was probably a
greater concentration of Arabic and Persian, and perhaps from XIII onwards also
Mongolian, loan-words in the literary than in the spoken language.

30. Possibly the earliest of these intermediate texts which, since it contains some
Mongolian words, cannot be earlier than XIII, is the didactic poem edited by
R. R. Arat under the title *Atebetü'l-hakayık*, Istanbul, 1951, cited as *At.* followed
by the lines.

31. An even earlier date has been suggested for a few Turkish *tafsīrs*, that is
interlinear translations of the Koran (mostly bilingual, some trilingual with a
Persian translation of the Arabic and a Turkish translation of the Persian) with
commentaries and other additional matter. The difficulty in determining the
date of these texts and the exact language in which they were composed lies in
the fact that the surviving manuscripts of them were all written later, perhaps
a good deal later, than the date at which they were composed and have been
subjected to a good deal of modernizing by the copyists, later phonetic forms,
and even actual words, being substituted for the author's own phonetic forms
and words. It is not at all clear whether all, or at any rate all the earlier, bilingual
manuscripts go back to a single archetype or whether some are independent
translations and compilations, but the latter is the more probable. Borovkov in
Tef. has indexed the Turkish vocabulary of one such work, the manuscript of
which was found in the town of Karshi. The name of the author, the date of
composition, the name of the copyist, and the date of copying are all equally
unknown, but the manuscript seems to be later than that of a similar *tafsir* found
in Turkey dated A.D. 1333 (see *Tef.*, p. 4) which is said to have a text close to that
of the Karshi manuscript. An examination of the vocabulary shows that it is
very heterogeneous. It includes a number of words of great antiquity for some
of which, e.g. **2 ap**, it is the latest authority. On the other hand, it also includes
Arabic, Persian, and even some Mongolian, loan-words, which proves that in its
present form it cannot be earlier than XIII, and it even contains different forms
of the same word, e.g. **aḍrıl-, ayrıl-, ayrul-**, which belong to different periods
in the history of the language. It has been suggested that these are evidence that
the language used, even if basically Xak., has been heavily influenced by Oğuz
and Kıp. and that it must therefore have been composed in some northern area
and ought to be classified as Xwarazmian (see para. 37). There is not, however,
any conclusive evidence that this is so, and there are other possible explanations
of the non-Xak. elements in the text. In the circumstances it seems better to
regard the language as intermediate between Xak. and Çağ., to date *Tef.* as
XIII(?) and cite it immediately after *At.*

32. Next in time, and in much the same language, comes the *Qiṣaṣu'l-anbiyā* of
Nāṣiru'l-dīn al-Rabğūzi composed in A.D. 1310. In this case, too, all the surviving

manuscripts are later than the work itself and have been subjected to a good deal
of modernization. No index of the vocabulary has been compiled, but a number of
words from a XIX printed edition are cited in Radloff's *Wörterbuch* and some of
these citations have been included here as XIV *Rbğ.*, followed by a reference to
Radloff.

33. One Arabic–Turkish vocabulary, the *Ḥilyatu'-insān wa ḥalbatu'l-lisān* of
Camālu'l-dīn ibnu'l-Muhannā (see *Studies*, p. 193) is probably of about the same
date as *Rbğ*. It contains an extensive list of Arabic words with the Turkish
equivalents. The author in his preface says that they are in two languages 'those
of our country and of Turkistan', and enumerates the phonetic differences
between the two. The facts given suggest that the first is an early form of Azer-
bayjani and the second a late form of Xak., but, except in a minimum number
of cases, it is impossible to decide whether any given word belongs to one of
these languages or both. Very probably the great bulk of them are early Azerbay-
jani, but as a matter of convenience it is easier to cite this work as XIV *Muh.*
immediately after *Rbğ*. followed by the page and line of Melioranski's edition
(*Mel.*) and the page of Kilisli Rif'at's edition (*Rif.*); the two editions are not
identical and it is likely that the latter, which is based on a single manuscript,
contains a good many words added to the author's text by later copyists.

34. Finally there are many texts in Çağatay, the earliest perhaps late XIV, the
great bulk XV or even a little later. The language still survives in a later form as
modern Uzbek, and no attempt has ever been made to define the latest date at
which Çağ. proper was still in use. Apart from a very extensive literature, three
major Çağ.–Persian and Çağ.–Old Osmanli dictionaries have been published
which, although they all profess to be primarily dictionaries of the language
used by Mīr 'Alī Şīr Nawā'ī, probably give a fairly exhaustive list of the Çağ.
vocabulary, excluding the Arabic and Persian, but not the Mongolian, words
which it includes. These are, in order of age:

(1) *Badā'i'u'l-luğat* written in Herat by Tāli'-i Harawī during the reign of the
Timurid Sultān Ḥusayn (A.D. 1438–1506), a Çağ.–Persian dictionary of which
the only(?) surviving manuscript, dated A.H. 1117 (A.D. 1705–6) was published
in facsimile with a comprehensive index by A. K. Borovkov, Moscow, 1961.

(2) The anonymous Çağ.–Old Osmanli dictionary commonly known (after the
first entry in it) as the *Abuşka*, compiled during the first half of XVI and published
by V. de Véliaminof-Zernof as *Dictionnaire djagatai-turc*, St. Petersburg, 1869.

(3) The *Sanglax*, a Çağ.–Persian dictionary written by Muḥammad Mahdī
Xān and finished in A.H. 1172 or 1173 (roughly A.D. 1759). A facsimile of one of
the best manuscripts was published with an introduction and comprehensive
indices by myself as E. J. W. Gibb Memorial, New Series XX, London, 1960.

Of these the *Sanglax*, although the latest, is much the most extensive and best,
the author correcting a good many mistakes made in previous dictionaries,
including the two mentioned above and others which have not survived. I have

cited the first occasionally as *Bad.* followed by the page, the second more frequently as *Vel.* followed by the page and the last, almost comprehensively, as *San.* followed by the page and line of the facsimile.

35. Exaggerated importance has been attached by some scholars to Şayx Sulayman Buxarī, *Luğat-i Çağatay wa Turki 'Oṭmānī*, Istanbul, A.H. 1298 (A.D. 1881); it contains very little original matter, apart from the mistakes which abound in it (see *San.*, p. 31) but I have occasionally cited it as *ŞS.* followed by the page.

36. There is in Bokhara an important manuscript of the *Muqaddimatu'l-Adab* of Maḥmūdu'l-Zamaxşarī, a classified Arabic glossary compiled early in XII with an interlinear translation in Turkish (as well as in Persian and Mongolian). The manuscript is dated A.D. 1495 and there is no prima facie reason for supposing that the Turkish translation is much older. The language is therefore likely to be Çağ. and this is confirmed by an examination of the Turkish equivalents in N. Poppe's index to the Mongolian translation in *Mongol'skii slovar' Mukaddimat al-Adab*, Moscow, 1938. Prof. Borovkov, when he died in 1963, had almost completed an index of the Turkish words in this manuscript, but it is uncertain whether it will now be completed and published. The Turkish entries in Prof. Poppe's work are not in a convenient form for use in preparing a Turkish dictionary, and are not, according to Prof. Borovkov, wholly accurate. I have not, therefore, quoted this work except occasionally (as *Zam.*).

37. The history of the northern stream of literary languages is a great deal more complicated. There was undoubtedly in XII a second literary focus further north than the area described in para. 28, centred perhaps in the court of the Xwārazmşāhs who first became important early in XII. The most convenient name for texts belonging to this stream is that recently adopted in such works as the *Fundamenta*, Xwarazmian (*Xwar.*), although the literary language which evolved there must also have been in use over a much wider area extending into southern Russia. The spoken languages in this area at this period were markedly different from Xak.; the Turks in Xwarazm and the Aral Sea area were Oğuz and those in southern Russia Kıpçak and already in XI Kāşğarī regarded Oğuz and Kıp. as separate languages fairly close to one another but different from Xak. Thus, so far as literary works are concerned, it is safe to assume that Xak. was very quickly modified to bring it closer to the local languages.

38. The oldest text from this area, if its colophon is to be relied on, was in fact composed not in Xwārazm but in the Crimea. It is the *Qiṣṣa-i Yūsuf* written by a certain 'Alī, of whom nothing further is known, in A.H. 630 (A.D. 1233). As usual the manuscripts are a good deal later and have been more or less modernized. D. Brockelmann listed a number of words from this work in *'Alī's Qissa-i Yūsuf, der älteste Vorläufer der osmanischen Literatur* (AKPAW, 1917). It has long been recognized that the language is not Osmanli and I have cited the words quoted in

this paper as Xwar. XIII '*Ali* followed by the page, but it may well contain more Oğuz elements than most other Xwar. texts.

39. Another text which can be tentatively identified as coming from the Aral Sea area is the enigmatic text sometimes called *Oğuz Nāme* (see *Studies*, pp. 48, etc.) written in the Mongolian Official Alphabet and edited by W. Bang and G. R. Rachmati, *Die Legende von Oghuz Qaghan* (SPAW, 1932). It contains several Mongolian loan-words and so cannot be earlier than XIII, but can hardly be much later. Considering its subject-matter, the language is likely to be much more specifically Oğuz than most other contemporary texts, but the ambiguity of the alphabet used makes it impossible to determine whether it has such a specifically Oğuz trait as initial **d-**. It is cited as XIII(?) *Oğ.*, followed by the line in the edition mentioned.

40. Next in order of time come five texts which are all more or less solidly dated. The oldest is the *Mu'īnu'l-murīd* written in Xwārazm in A.D. 1313 (see Eckmann, op. cit., p. 115), but I have not been in a position to cite any words from it.

41. The next is the *Xusraw u Şīrīn*, written by a certain Quṭb at the court of Tini Beg Xan of the Golden Horde in A.D. 1341–2. It has been admirably edited by A. Zajączkowski, *Najstarsza wersja turecka ḫusräv u Širin*, 3 vols., Warsaw, 1958–61, with a facsimile of the only manuscript, a transcription, and a full index excluding the Arabic and Persian loan-words. It is cited as XIV *Quṭb* followed by the page in Zajączkowski's index.

42. The third is the *Muḥabbat Nāma* of Xwarazmi. Two manuscripts of this poem have survived, the second containing a number of verses which did not form part of the original text; see my paper on this poem in *CAJ* VII. 4, 1962. Unfortunately verses 437 and 440 which state that the poem was composed 'on the banks of the Sir Darya in A.H. 754 (A.D. 1353)', are among the interpolated verses but the date must be approximately correct, since it falls within the reign of Cani Beg Xan of the Golden Horde, during whose reign the poem was composed. It is cited from the recent editions (see *Studies*, p. 48) as XIV *MN* followed by the number of the verse.

43. Next there is the *Nahcu'l-Farādīs* written by Maḥmūd ibn 'Alī in Xwārazm, or perhaps Saray on the Volga, sometime before A.D. 1358, the date of the earliest manuscript, but not necessarily much before, since there is fairly good evidence that the author did not die until A.D. 1360. A reproduction of one of the best manuscripts was published in J. Eckmann, *Nehcü'l-Feradis* I, Ankara, 1956, but unfortunately his index is not yet published.

44. Finally there is the free translation into Turkish of Sa'dī's *Gulistān*. Only one manuscript of this work has survived, and a facsimile of it, with a preface by

Prof. F. N. Uzluk, was published by the Türk Dil Kurumu in 1954 under the title *Seyfi Serâyî Gülistan Tercümesi*. The translator was a native of Sarai on the Volga, but made his translation at the Mamluk court in Egypt in A.D. 1391.

45. Quite recently there has appeared the first half of an admirable dictionary by E. Fazylov of *Qutb, MN*, the *Nahcu'l-Farādis*, and the *Gulistan*, with extensive quotations and references, under the title *Starouzbekskii yazik. Khorezmiiskie pamyatniki XIV veka*, Tashkent, 1966. I have used this work extensively to provide references to the *Nahcu'l-Farādis*, under the title *Nahc.* followed by the page and line in Eckmann's facsimile for those words contained in this part of the dictionary; for the rest of the vocabulary I have had to rely on a very inadequate list of words published by Kivâmettin in *TM IV*. I have not thought it necessary to add references to the Gulistan (*Gul.*) except very occasionally, since its vocabulary is practically identical with that of the other works mentioned.

46. Apart from the texts mentioned above a few other Xwar. texts have survived and one or two short ones have been published, at any rate in facsimile, but are not cited herein.

47. Seven vocabularies of medieval Kıp. have been published, some of them containing some Türkmen (*Tkm.*) words. The most important, and the only one which is purely Kıp., is the Codex Cumanicus, a handbook of the Koman (Kom.) language, a Kıp. dialect, in Latin, compiled early in XIV (see *Studies*, p. 48 and *Fundamenta*). The text is not entirely homogeneous; the earlier part was compiled for Italian merchants, but it also contains later additions by German missionaries. An index to it was published in K. Grønbech, *Komanisches Wörterbuch*, Copenhagen, 1942, and is cited as XIV *Kom. CCI* (for the Italian) and *CCG* (for the German part); *Gr.* sometimes followed by the page.

48. The remaining six are all in Arabic, one a list of Turkish words in alphabetical order with Arabic equivalents, the others Arabic handbooks with Turkish equivalents. All these vocabularies have a specified or inferable Egyptian Mamluk background, and it seems clear that basically they are handbooks of the languages spoken by Turkish slaves brought to Egypt from southern Russia, some of them spoke Kıp. and others Tkm., an Oğuz dialect. For example, *Hou.* (see para. 49) seems from internal evidence to have been compiled from oral information collected in the Mamluk possessions in Syria. In some cases an Arabic word is translated by different Kıp. and Tkm. words, in others individual words are described as Kıp. or Tkm., but in the great majority of cases no such indication is given. The presumption in nearly all cases is that the word is Kıp., but there are one or two passages in *Id.* (see para. 50) which seem to imply the contrary.

49. Until almost today it was generally accepted that the oldest of these vocabularies was the anonymous Arabic–Turkish vocabulary in Leiden manuscript

517, Warner, published in M. T. Houtsma, *Ein türkisch-arabisches Glossar*, Leiden, 1894, cited herein as XIII *Hou.* followed by the page and line of the Arabic text. But as this book was actually going to press an article by Barbara Flemming (Hamburg) was published in *Der Islam*, Band 44, Berlin, June 1968, pp. 226 ff., in which she announced that Houtsma, who purported to publish the exact text of this unique manuscript had converted the date in the colophon into the figures 643, whereas the facsimile published in her article shows quite clearly the words *talāta wa arba'in wa sab'umi'ya* that is (Şa'bān) A.H. 743 equivalent to (January) A.D. 1343. Instead, therefore, of dating back to XIII and being the oldest of these vocabularies, it goes back only to the middle of XIV and comes chronologically between the two vocabularies mentioned in the following paragraphs. This announcement unfortunately appeared too late for it to be possible to correct the numerous references to *Hou.* in the dictionary.

50. There are two other XIV vocabularies. The *Kitābu'l-idrāk li-lisāni'l-atrāk*, the only list of Turkish words with Arabic equivalents, was written in Egypt, probably in A.D. 1313, by Aṭīru'l-dīn Abū Ḥayyān Muḥammad ibn Yūsuf, originally a native of Granada. A critical edition was published by A. Caferoğlu, *Abu Hayyân, Kitâb al-İdrâk li-lisân al-Atrâk*, Istanbul, 1931. It is cited as XIV *İd.* followed by the page in the Arabic text.

51. The *Kitāb Bulġati'l-muştāq fi luġati'l-turk wa'l-qifçāq* was written by Camālu'l-din Abū Muḥammad 'Abdullah al-Turki probably in the middle of XIV and perhaps also in the Mamluk dominions in Syria. The only manuscript, which is not quite complete, was published with a comprehensive index by A. Zajączkowski, *Manuel arabe de la langue des Turks et des Kiptchaks*, Warsaw, I, 1938; II (title in Polish), 1954. It is cited as XIV *Bul.* followed by the page and line of the printed text in I and the page of the manuscript in II (verbs).

52. There are two XV vocabularies, both surviving in unique manuscripts and dating from early in the century, but it is uncertain which is the older. *Al--tuhfatu'l-zakiya fi luġati'l-turkiya*, an Arabic–Turkish vocabulary in alphabetical order, mainly Kıp. with a number of Tkm. words, was written almost certainly in Egypt and before A.D. 1426 (the date of a note on the first page). It was edited with a facsimile, translation, and index in B. Atalay, *Ettuhfet-üz-zekiyye fil--lûgat-it-türkiyye*, Istanbul, 1945. It is cited as *Tuh.* followed by the page and line of the facsimile.

53. *Al-qawāninu'l-kulliya fi-ḍabţi'l-luġati'l-turkiya* was compiled by an unknown author in Cairo by cross-questioning Turks living there. The book is undated, but a reference in it to Tamerlane makes it certain that it was written in XV, and probably early in the century. The text was published by Köprülüzade Mehmed Fuad (Prof. F. Köprülü), Istanbul, 1928. It is cited as *Kav.* followed by the page and line.

54. Quite recently Prof. Zajączkowski published an account of a short Arabic–Kıp. vocabulary entitled *al-Durratu'l-muḍi'a fi luğati'l-turkīya* discovered in Florence, with an extract from the text and an index to the extract, in *Rocznik Orientalistyczny* XXIX, Pt. i, pp. 39 ff. The work is anonymous and undated but clearly belongs to the first half of xv. It adds little to our knowledge of Kıp. but I have quoted a few words from it as xv *Dur.* followed by the page.

55. The Oğuz language which is conveniently, but not quite accurately, called 'Old Osmanli' (*Osm.*), since the earliest texts in it date from before the foundation of the Ottoman Empire, is recorded from mid xiii, but the xiii material is scanty. The earliest is a few verses in the works of Mawlānā Calālu'l-din al-Rūmi (A.D. 1207/8–1273/4) and a few verses in those of his son commonly called Sultan Veled. These verses might more properly be classified as Salcuk, but as they are the first stage of a continuous literary tradition it seems more convenient to classify these very early texts and other slightly later texts which are not strictly Osm., like the *Kitāb Dede Korkut*, which seems to be in a Tkm. language, probably xv., as *Osm.* so as to avoid an excessively complicated terminology. A good many early works of this kind have been published with partial or complete indices, but it did not seem to me necessary for the purposes of this dictionary to do more than cite under the heading xiv ff. *Osm.* (since the xiii material is so inconsiderable) those of the words contained in this dictionary which reappear in the Türk Dil Kurumu publication *Tanıklariyle Tarama Sözlüğü*, 4 vols., Istanbul, 1943–5; Ankara, 1953–7 (*TTS I–IV*) or as '*Rūmi*' words in *San*. A new edition of *TTS* is now being published, but is not yet sufficiently far advanced to make citation from it very convenient.

THE MODERN LANGUAGES

56. It would not be possible to include in this book all the evidence which could be assembled regarding the survival in modern languages of the early words listed herein without a much greater delay in its publication and a much greater increase in its bulk than would be regarded as advantageous. In any event it is sufficiently obvious to all students of the Turkish languages that common words like at 'a horse' and al- 'to take' still survive almost everywhere to make a detailed proof of the point unnecessary. Nor is it very important, when a basic word is known to survive in some or all modern languages, to know whether its various derivatives survive also. On the other hand, for the reasons stated in para. 8, the subject cannot be entirely neglected, and it is important to include at any rate some references to modern languages. This immediately raises the question how modern languages should be classified and grouped. A good many systems of classification have already been suggested; there are two quite different ones in *Fundamenta*; but for present purposes it is hardly necessary to use anything more elaborate than a simple geographical system. Even this is open to some objections, since there are no clear-cut geographical lines between the modern descendants of the early languages, and some modern languages

descended from different ancient languages have in the course of time de-
veloped characteristics which bring them closer to one another than to languages
with which they are genetically related. After a good deal of reflection I finally
decided upon a sixfold division: north-eastern (NE), south-eastern (SE), north-
central (NC), south-central (SC), north-western (NW), and south-western
(SW). Of these the NW languages are, genetically speaking, Kıp. and the SW
Oğuz; the SC and probably SE are, broadly speaking, descended from Xak.
but by different lines of descent.

57. In the NE group are included the languages spoken, but until recently not
written, in eastern Siberia and adjacent areas. Yakut, however, which has been
isolated from the rest so long that it has acquired very special characteristics of
its own, is seldom cited, the authority used being E. K. Pekarsky, *Slovar'
yakutskogo yazyka*, St. Petersburg, 1907–30, cited as *Pek.* The most important
authority for the vocabulary of these languages is W. Radloff, *Versuch eines
Wörterbuches der türk-Dialecte (Opyt slovarya tyurkskikh narechii)*, 4 vols.,
St. Petersburg, 1888–1911, cited as *R* followed by the volume in roman, and the
column in arabic numerals, individual languages being mentioned in the abbre-
viated forms used in that work with the minor alterations of spelling shown in
the List of Abbreviations. I have also used the *Khakassko-russkii slovar'*, edited
by N. A. Baskakov, Moscow, 1953, cited as *Khak.* sometimes followed by the
page, the *Tuvinsko-russkii slovar'*, edited by A. A. Pal'mbakh, Moscow, 1955,
cited as *Tuv.* sometimes followed by the page, and occasionally the modern
Russian–Khakas, Russian–Tuvan, and Russian–Altay dictionaries.

58. In the SE group are included the Turkish languages and dialects of Chinese
Turkistan and adjacent areas, which have traditionally been called Eastern
Türki and are now called Neo-Uyğur by Soviet and Chinese scholars. These
fall into two main groups, the literary language written in Arabic script and the
spoken dialects. I have called both *Türki.* For the first, citations are from R. B.
Shaw, *A Sketch of the Turki Language*, Calcutta, 1878, cited as *Shaw* followed
by the page, and Burhan Şehidi, *Uyğurça-Xenuça-Rusça Luğat (Uigursko-
kitaisko-russkii slovar')*, Pekin, 1953, cited as *BŞ* followed by the page.
A good many words in one dialect, Tarançı, spoken in southern Siberia are
included in *R.*, and are cited as *Tar.* followed by a reference to *R.* For the rest,
mainly dialects spoken in southern Sinkiang, I have used G. Jarring, *An Eastern
Turki-English Dialect Dictionary*, Lund, 1964, cited as *Jarring* followed by the
page. One language in this area is in a class by itself, that of the Sarığ Yuğur
in Kansu, the only language which can reasonably be regarded as directly
descended from early Uyğ. It is recorded in S. Ye. Malov, *Yazyk zheltykh
uigurov*, Alma Ata, 1957, cited occasionally as *Sarığ Yuğ.* followed by the page.

59. In the NC group are included Kırğız (Kır.) and Kazax (Kzx.), called in *R*
Kara Kırğız and Kırğız respectively. Historically Kır. belongs to the same family
as the NE languages and Kzx. seems to be a Kıp. dialect, but the peoples talking

them have lived in close propinquity to one another for so long that they now
have many common characteristics. The authorities which I have used princi-
pally are, for Kır., K. K. Yudakhin's *Kirgizsko-russkii slovar'*, Moscow, 1965,
cited as *Yud.* followed by the page, and for Kzx. Kh. Makhmudov and G. Musa-
baev, *Kazakhsko-russkii slovar'*, Alma Ata, 1954, and B. N. Shnitnikov's
Kazakh-English Dictionary, The Hague 1966, cited as *MM* and *Shnit.* respec-
tively followed by the page. I have also occasionally used the equivalent Russian–
Kır. and Russian–Kzx. dictionaries, and only occasionally quoted *R* since the
modern authorities seem to be rather fuller and more reliable.

60. There are several modern SC spoken dialects, but the only literary language
is Uzbek (Uzb.) for which I have used A. K. Borovkov, *Uzbeksko-russkii
slovar'*, Moscow, 1959, cited as *Bor.* followed by the page, and less often the
Russian–Uzb. dictionary of 1954. The vocabulary in H. Vambery's *Čagataische
Sprachstudien*, Leipzig, 1867, is very heterogeneous in character. The bulk of
the words in it are classical Çağ. taken, often with some errors, from various
oriental authors; the remainder seem to be colloquial XIX Uzb. words collected
by the author himself, and occasionally cited as *Vam.* followed by the page.

61. The NW group comprises a rather wide range of languages. Those quoted
in *R* include Kazan Tatar (*Kaz.*) and three Karaim dialects, those of the Crimea,
Lutsk, and Troki (*Kar., Krm., L.,* and *T.*) but he also uses *Krm.* for the non-
Karaim language spoken in the Crimea which is indistinguishable from Osm.,
so that the only safe indication that a word is Karaim is that it is given in the
Hebrew alphabet. The words quoted from these languages are reproduced with
the same abbreviations and the reference in *R*. For Kar. T. I have occasion-
ally used T. Kowalski, *Karaimische Texte im Dialekt von Troki*, Cracow, 1929,
cited as *Kow.* followed by the page. For (Kazan) Tatar I have used the *Tatarsko-
russkii slovar'*, Moscow, 1966, cited as *Kaz.* or *Tat.*, and for Karakalpak (*Kk.*)
and Nogay (*Nog.*) I have used the *Karakalpaksko-russkii slovar'*, Moscow,
1958 and the *Nogaisko-russkii slovar'*, Moscow, 1963, both edited by N. A.
Baskakov and cited by the page, and occasionally the Russian–Kk. and –Nog.
dictionaries. For Kumyk (*Kum.*) there is at present available only the *Russko-
kumykskii slovar'*, Moscow, 1960, edited by Z. Z. Bammatov; Bashkir has
diverged so far from the rest of the group that I have not taken it into account.
No convenient authorities for minor languages like Karaçay and Balkar are at
present available. As regards Chuvash (*Çuv.*) see para. 2.

62. In the SW group are included only three languages, Azerbayjani (Az.), Osm.
(including Rep. Turkish), and Tkm. For Az. I have used various authorities
published in the Soviet Union, not citing them individually. There are many
authorities for Osm., but those which I have generally used are Sir James
Redhouse, *A Turkish and English Lexicon*, in the Constantinople, 1921, reprint,
cited as *Red.* followed by the page; Şayx Sami, *Qāmūs-i Turkī*, 2 vols., Istanbul,
A.H. 1318 (A.D. 1900–1) cited as *Sami* followed by the page; and occasionally the

A.H. 1306 (A.D. 1888) edition of the *Lehce-i Osmaniya*, cited as *Leh.* followed by the page. In noting the survival of words in modern languages I have not included those words which have recently been re-introduced into Rep. Turkish to replace Arabic and Persian loan-words. Some of these are genuine old words which had become obsolete in Osm., others are old loan-words like **acun** (for a:ju:n, a word borrowed from Sogdian) and **çağ** 'period' (a medieval loan-word from Mong.), but they have not had a continuous history in Osm.

63. As pointed out in paras. 2 and 12 it is often important in tracing the history of a Turkish word to know whether it became a loan-word in Mongolian and, if so, when, since this often throws light not only on the age of a word but also on its original pronunciation and meaning. The question of such loan-words is discussed at length in *Studies*, Chap. 11, and it is there suggested that these words were borrowed in three main periods, the first, probably v and vi, during which words were borrowed by a Mongolian-speaking people, possibly the Kitañ, from an l/r Turkish-speaking tribe, possibly the Tavğaç; the second, probably between viii and xii; and the third during and after the reign of Chinggis. The earliest substantial remains of Mongolian are not, however, earlier than xiii. In citing Mongolian words borrowed from Turkish I have normally used three authorities:

(1) E. Haenisch, *Wörterbuch zu Manghol un Ni'uca Tobca'an*, Leipzig, 1939, which contains a list of the words in the well-known Mong. xiii *Secret History* (SH), cited as *Haenisch* followed by the page.

(2) The standard, though now somewhat antiquated, dictionary of Classical Mongolian, J. E. Kowalewski, *Dictionnaire mongol-russe-français*, Kazan, 1841 ff., cited as *Kow.* followed by the page.

(3) A good dictionary of modern written Mongolian, *The Mongol–English Practical Dictionary*, compiled by M. Haltod and others and published by the Evangelican Alliance Mission (in the U.S.A.) 1949–53, cited as *Haltod* followed by the page.

OTHER MODERN AUTHORITIES

64. In addition to the modern authorities quoted above use has been made of the following other modern authorities:

(1) A. Caferoğlu, *Uygur Sözlüğü*, 3 parts, Istanbul, 1934–8, a useful list of the words contained in the indices to the texts enumerated in paras. 20 (2) (e) and 21 above and one or two others, cited as *Caf.*

(2) v. G., *ATG*, see para. 20 (2) (e).

(3) G. Doerfer, *Türkische und mongolische Elemente im Neupersischen*; I *Mongolische Elemente*, Wiesbaden, 1963; II *Türkische Elemente, alif bis tā*, 1965; III ditto *jīm bis kāf* 1967, cited as *Doerfer*, followed by the volume and the number of the word.

(4) C. Brockelmann, *Osttürkische Grammatik der islamischen Litteratur-sprachen Mittelasiens*, Leiden, 1954, cited as *Brockelmann* followed by the number of the paragraph.

(5) A. M. Shcherbak, 'Nazvaniya domashnikh i dikikh zhivotnykh v tyurk-skikh yazykakh' ('The names of domesticated and wild animals in the Turkish languages'), one of several articles in *Istoricheskoe razvitie leksiki tyurkskikh yazykov*, Moscow, 1961, cited as *Shcherbak*, followed by the page.

(6) The reproductions of MS. No. S. J. Kr. 4638, Packet 8 in the manuscript collection of the Leningrad branch of the Institute of the Peoples of Asia, recently published in D. I. Tikhonov, *Khozyaistvo i obshchestvennyi stroi uigurskogo gosudarstva X—XIV vv.*, Leningrad, 1966, and again with transcription and translation by E. R. Tenishev in an article called 'Khozyaistvennye zapisi na drevneuigurskom yazyke' in *Issledovaniya po grammatike i leksike tyurkskikh yazykov*, Tashkent, 1966. The manuscript is a late Uyğ. family archive, and is cited as *Fam. Arch.* followed by the line.

THE ARRANGEMENT OF INDIVIDUAL ENTRIES

65. In paragraphs 4, 5, and 8 some account has been given of the kind of information which is supplied about individual words. There is not a single word regarding which the full range of information is available, and it may therefore be useful to set out here what the full range is and how the various items are arranged. The word may be preceded by a code letter (see paragraph 18). If it is a *hapax legomenon* (see paragraph 7) it is immediately followed by 'Hap. leg.'. From this point the order varies slightly. It is sometimes convenient to enter next the original meaning possibly with some remarks on later developments, but this is superfluous if the word is Hap. leg., since the quotation will supply it, and if the word is derived it is more logical to record the derivation before the meaning. After this it may be desirable to add some other remarks. If the fact that the word is a loan-word in Mongolian is relevant, this is mentioned next. Information regarding the survival of the word then follows. It cannot be given earlier since there are many instances of words surviving in modern languages, usually NE, not because they have existed continuously in those languages but because they have been reborrowed from Mongolian, perhaps quite recently. Finally, cross-references may be given to other words etymologically related or other words with the same or a similar meaning. The quotations then follow in the following order: **Türkü** VIII; VIII ff.; VIII ff. Man.; VIII ff. Yen.: **Uyğ.** VIII; IX; VIII ff. Chr.; Man.-A; Man.; Bud.; Civ.; XIV *Chin.-Uyğ. Dict.*: **O. Kır.** IX ff.: **Xak.** XI (including *KB*); XII(?) *KBVP*; XIII(?) *KBPP*; *At.*; *Tef.*; XIV *Rbğ.*, *Muh.*: **Arğu, Çiğil, Gancak** XI: **Çağ.** XV ff.: **Oğuz** XI: **Xwar.** XIII *'Ali*; XIII(?) *Oğ.*; XIV *Quṭb*, *MN*, *Nahc.*: **Kıp.** XI: **Kom.** XIV *CCI, CCG*; *Gr.*: **Kıp./Tkm.** XIII *Hou.*; XIV *Id., Bul.*; XV *Tuh., Kav., Dur.*: **Osm.** XIV ff.

66. In conclusion may I very humbly submit that this book contains a vast number of quotations, translations, and references, and that it is hard to believe

that in such a large flock there are no black sheep? May I plead, as the father of Turkish lexicography did nine centuries ago, that **yazma:s atım yağmur yaṇılma:s bilge: yaṇku:** 'the only shot that never misses is the rain, the only scholar who never makes a mistake is the echo' *Kaş. III* 379, 20, and present in advance my apologies for these and other shortcomings.

It would not be proper for me to end this preface without expressing my sincere gratitude to the British Academy for their generous contribution towards the cost of producing this volume.

GERARD CLAUSON

August 1968