

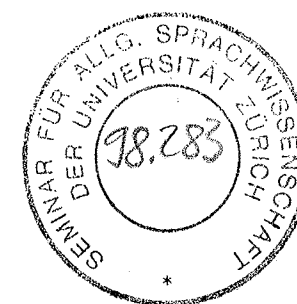
Nb 17 556

**Other works in the series**

The Romance Languages  
The Celtic Languages  
The Slavonic Languages  
The Germanic Languages  
The Semitic Languages  
The Indo-European Languages  
The Dravidian Languages  
The Uralic Languages

# THE TURKIC LANGUAGES

EDITED BY  
Lars Johanson  
and  
Éva Á. Csató



LONDON AND NEW YORK

First published 1998  
by Routledge  
11 New Fetter Lane, London EC4P 4EE  
29 West 35th Street, New York, NY 10001

Selection and editorial matter © 1998 General editors  
The chapters © 1998 Routledge

Typeset in 10.5/12 Times by Solidus (Bristol) Ltd, Bristol  
Printed and bound in Great Britain by MPG Books Ltd, Bodmin

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*British Library Cataloguing in Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging-in-Publication Data*

A catalog record for this book is available on request

ISBN 0-415-08200-5

---

# Contents

<b>List of Maps and Figures</b>	viii
<b>List of Tables</b>	ix
<b>List of Contributors</b>	xii
<b>Preface and Acknowledgements</b>	xiii
<b>The Turkic Languages (map)</b>	xvi
<b>Notes on Transcription and Symbols</b>	xviii
<b>List of Abbreviations of Grammatical Terms</b>	xxiii
<b>1 The Speakers of Turkic Languages</b> Hendrik Boeschoten	1
<b>2 The Turkic Peoples: A Historical Sketch</b> Peter B. Golden	16
<b>3 The Structure of Turkic</b> Lars Johanson	30
<b>4 The Reconstruction of Proto-Turkic and the Genetic Question</b> András Róna-Tas	67
<b>5 The History of Turkic</b> Lars Johanson	81
<b>6 Turkic Writing Systems</b> András Róna-Tas	126
<b>7 Old Turkic</b> Marcel Erdal	138
<b>8 Middle Kipchak</b> Árpád Berta	158

# Notes on Transcription and Symbols

## Transcription

The official orthography of the Republic of Turkey is used for Modern Turkish. For languages spoken outside Turkey, a variety of alphabets and orthographical conventions are applied, which makes transcription necessary. The recently introduced Latin-based scripts for Azerbaijanian, Turkmen, etc. are not yet sufficiently established. A transliteration of the Cyrillic scripts, which differ considerably from each other, would distort the phonetic shape and make comparison between the languages difficult. All examples from languages which use the Cyrillic or Arabic script will thus be given in transcription. The system applied, referred to as the 'Turcological notation', is basically identical to the traditional 'Fundamenta' notation as represented in Deny *et al.* 1959, pp. xiv–xv (see References and Further Reading, pp. 122–5).

The sections 'Sound system(s)' primarily deal with the sounds of the language in question and relations among them. In spite of differing approaches to the problems, these sections generally contain remarks on the inventory of phonemes, i.e. which sounds can be used in the language to distinguish words. Though some authors tend more strongly towards a phonemic representation, the transcription will in general not only represent phonemes, but also major allophonic variants, thus conveying an idea of the actual pronunciation. For example, front *g* and *k* sounds will as a rule, irrespective of their phonemic status, be rendered as *g* and *k*, and their back equivalents as *ɣ*/*ġ* and *q*. The phonetic difference between a front *l* and a back *ɫ* will not, however, be marked. The transcription is thus relatively broad. Wherever necessary, finer phonetic details are given in the transcription of the International Phonetic Association (IPA).

## Vowel Notations

The transcription employed in the present volume differs from the 'Fundamenta' notation mainly with respect to the representation of *e*-sounds. Thus, the normal low type [e] with all its variants is written as *e* instead of the 'Fundamenta' notation *ä*. If a particular language also has a higher *e*

phoneme, a so-called 'closed *e*', it is rendered as *è*. If, in addition to *e*, a lower phoneme is present, it is written *æ*. The low front unrounded Uzbek vowel is transcribed as *ä*.

The sign used to denote a front unrounded high vowel is *i*. A corresponding back vowel is written as *ĩ* (in Turkish orthography *ı*). A front rounded high vowel is written as *ü*, a back rounded high vowel as *u*, a back rounded low vowel as *o*, and a back unrounded low vowel as *a*. A labialised variant of *a* is rendered by *â*, e.g. in Uzbek. Centralised, retracted or lowered variants of vowels are written with an under-dot, e.g. *ĩ*. Centralisation can also be denoted by an over-dot. Thus, in the transcription of Karaim words, *û*, *ô* are to be read as retracted *ü* and *ö* respectively. The weaker element of a diphthong is marked with a raised letter, e.g. <sup>i</sup>*a*, <sup>i</sup>*e*, <sup>u</sup>*o*, <sup>ü</sup>*ö*. Reduced vowels are written with the diacritic breve, e.g. *ă*, *ĕ*. The sign <sup>o</sup> is used to signal a reconstructed, probably reduced, vowel entity. Long vowels are indicated by a macron, e.g. *ā*, and semi-long vowels, e.g. in Khalaj, by a dot following the vowel, e.g. *a˙*.

## Consonant Notations

The following table includes the most important consonant signs occurring in the volume.

Table i Consonants

	Labial	Prepalatal	(Post)palatal	Velar, uvular, glottal
stop	p, b	t, d	k, g	q, ġ, ʔ
fricative	f, v, θ, ð	s, z	š [ʃ], ž [ʒ]	ɣ, ɣ̣, h
nasal	m	n	ɲ	ŋ
affricate		c [ts], dz	č [tʃ], ĵ [dʒ]	
glide	w		y	
liquid		l, ɫ, r		

In the above table, some IPA correspondences are given in square brackets []. In some chapters, the term velar is used to cover both postpalatal and velar/uvular dorsals.

Palatalisation is marked with an apostrophe, as in *k'*. In some cases, the diacritic is placed over the consonant sign; thus, the palatalised *n* and *s* are written *ɲ* and *š*, respectively. Reduced voice is denoted with an under-ring, e.g. *ḥ*. Some authors use small capitals for lenis variants with reduced voice, e.g. Chuvash [ebir] 'we' = *eb̌ir* (p. 435). Aspiration is indicated with a superscript *h*, e.g. *p<sup>h</sup>*. The acute accent sign marks high pitch accent, e.g. *á*, whereas the vertical stroke ' indicates stress accent on the following syllable. Deviating notations are explained in the relevant chapters.

overall upper chronological limit, the end of the Mesolithic Period or rather the beginning of the Neolithic Period. The changes connected with the Neolithic Revolution created the sociolinguistic conditions for a system of communication that led to the type of languages we know. In the region concerning us here, the Neolithic Period started at about 4500–4000 BC. If Proto-Turkic reflects a language of the first period, it began around this time. If there was an earlier proto-language, e.g. a language in the sense of a reconstructed Proto-Altaic, Proto-Turkic came into existence only after its dissolution. The lower limit of Proto-Turkic is the time of the appearance of the first direct data from existing Turkic languages, in fact after the separation of the branches of Turkic, i.e. about the middle of the first millennium BC.

Proto-Turkic must be conceived of as a successive set of synchronous systems in the Ancient Turkic period, a period from which we have no direct data, but which we can reconstruct by comparing existing Turkic languages and their historical documents. The Proto-Turkic we can reconstruct best is the layer nearest to our sources. We denote this system as Proto-Turkic<sub>0</sub>, or simply Proto-Turkic, and earlier systems as Proto-Turkic<sub>-1</sub>, Proto-Turkic<sub>-2</sub> etc. As for Ancient Turkic, it can be divided into early Ancient Turkic, the period in which no stabilised dialects existed, and late Ancient Turkic, when the dialect boundaries were stabilised.

#### *The Question of the Original Homeland*

The term *urheimat* 'original homeland' denotes the region where a language has come into being, in our case the place of the formation of Turkic. However, it is not necessary to suppose that the ancestors of the Turks remained at the place where Turkic had been formed up to the dissolution of Ancient Turkic. If the entire Turkic-speaking group moved from there to relatively distant areas, a second, a third and even more *urheimats* can be implied.

There are several methods of reconstructing the place of an *urheimat*. The lexical stock of the proto-language can give hints through words denoting parts of the natural environment peculiar to a region: flora, fauna, climate. Early loanwords from other (proto-)languages may be helpful, if the geographical setting of these languages is known. Borrowing mostly implies language contact, which means historical contacts among peoples. Archaeologists try to trace archaeological cultures of identified speaker groups back to earlier groups. Such methods have their limitations and must be used with great caution. Even then, the only *urheimat* we can determine is the last one, the place where the Turks lived before the dissolution of the Ancient Turkic unity. The last habitat we can reconstruct with our data and existing methods can be placed in west and central Siberia and in the region south of it.

#### *Early Dialects and Isoglosses*

Though Ancient Turkic must be considered as a successive set of systems of one language, this unity was not a homogeneous one. In early Ancient Turkic,

from about 3000 until 500 BC, no stabilised dialects can be assumed. In late Ancient Turkic, the 'ancestor dialects' of the later Turkic languages were formed. At the end of early Ancient Turkic, the formation of the isoglosses *r/l* and *z/š* had begun. This means that, in special phonetic environments, the oppositions *r* vs. *z* and *l* vs. *š* were neutralised and new sounds appeared. Later on, the change was generalised areally to the effect that certain words were pronounced with *z* and *š* in one area and with *r* and *l* in another area, e.g. *\*buzagu* vs. *\*buragu* 'calf', *\*tāš* vs. *\*tāl* 'stone'. Following the split into different Turkic branches, one bunch of isoglosses was generalised in one branch and another bunch in the other. However, certain features existed in more than one branch. Thus, Proto-Turkic initial *\*y-* was replaced by *ǰ-* not only in the ancestor of the *r/l* languages, but also in some of the *š/z* languages. Furthermore, the lexical isoglosses of Ancient Turkic did not always follow the phonological ones. Thus the dialect distribution of the words *tūš* and *tūl* 'dream' did not coincide exactly with the isoglosses of *š* and *l*. As a result, *tūl* is found in Old Uyghur texts beside the normal form *tūš*.

### **A Sketch of Proto-Turkic**

Sources available for the reconstruction of Proto-Turkic include comparison of present-day Turkic languages, historical documents written in Turkic languages, loanwords copied by non-Turkic languages from Turkic languages and loanwords copied from non-Turkic languages by one or more Turkic languages.

While for a long time, the reconstruction of Proto-Turkic was mainly based on East Old Turkic and the Oghuz languages, a more thorough comparison of all known written and spoken languages began later on. Nevertheless, the Runic, Uyghur, Manichaean and other documents of East Old Turkic still play a key role in the reconstruction, since west Old Turkic languages such as Old Chuvash and Khazar are much less documented and thus difficult to reconstruct.

In the following outline of Proto-Turkic, special attention will be paid to features common to Chuvash and East Old Turkic. Some of them may be secondary, for instance, developed in Chuvash under Tatar influence, and some features that are absent may have been present in Proto-Turkic but lost later on. If these possibilities are kept in mind, the features common to East Old Turkic and Chuvash may serve as a solid basis for reconstructing Proto-Turkic.

### **Phonology**

#### *Vowels*

As for vowel quantity, Proto-Turkic had two distinctive vocalic lengths which were realised in first syllables as short (or 'normal') vs. long vowels. Non-first syllables had short and reduced vowels, but it is still unclear whether there

was an opposition between a 'normal' and a reduced quantity. For a possible third Proto-Turkic vowel quantity, see p. 90.

The known Turkic languages show no traces of early diphthongs, but rather segments in which one part is a vowel and the other a consonant *w* or *y*. Still we cannot exclude the possibility that Proto-Turkic had genuine diphthongs with two vowel elements. They were then probably of transitory character, i.e. emerging from long vowels or reflecting foreign diphthongs in loanwords, e.g. Old Turkic *tōn* 'clothing' < \**taunā* ← Indic *tauna*. Some interesting features in Khalaj can be attributed to old diphthongs (see pp. 93–5).

The shortening of long vowels was gradual and did not take place at the same time in all early Turkic languages. It is not even necessary to assume that all types of long vowels were shortened in all positions at the same time within one and the same language. Old Turkic exhibits a transitory stage in the process of shortening.

Theoretically, Proto-Turkic might have had four *e*-sounds, *e*, *ē*, *è*, *ē̄*. It is at least safe to assume that it had a short closed *è* beside an open *e*. It is, however, not necessary to assume a back vowel \**ē̄* in order to explain Common Turkic *a* ~ Chuvash *ī* in certain words (see p. 92). In some Turkic languages, e.g. those spoken in the Volga region, *a* was rounded. The *a*-sounds that remained unrounded later changed into *ī* in Chuvash, Yakut and Tuvan (in different words). There is no reason to reconstruct a rounded *ā* for Proto-Turkic.

In non-first syllables, Proto-Turkic did not only have *-u/-ū*, but most probably also *-o/-ö*. The high vowels *u*, *ū*, *i* and *ī* were reduced word-finally and disappeared in Old Turkic, whereas the half-low ones became high: *-o > -u*, *-ö > -ū*, *-è > -i*. The low *-e* was raised to *-è*. The low *-a* partly remained as such and was partly rounded to *-ā* or raised to *ī*.

In certain Old Turkic documents, *a* is written instead of an expected *ī* in non-first syllables. While some scholars have taken this notation to represent

Table 4.1 Vowels in Proto-Turkic first syllables

Short vowels				Long vowels					
ī	u		ū	ī	ū		ū	ī	
	o		ö		ō		ē		
	a	e			ā				

Table 4.2 Vowels in Proto-Turkic non-first syllables

Short vowels				Reduced vowels					
ī	u		ū	ī	ū		ū	ī	
	o		ö		ō		ē		
	a	e			ā	ē			

a special Old Turkic or even Proto-Turkic sound, it is probably a transcription of the still preserved back *ī*, mostly in the vicinity of *k*, *g*, *r*, *l* (cf. pp. 107–8).

It has also been claimed that traces of an old vowel gradation can be found in stem alternations such as *bi* 'I' ~ *bē-n-* (oblique stem). This is, however, a structurally conditioned sound change that seems more similar to an umlauting phenomenon.

### Consonants

Proto-Turkic had a fortis (strong) vs. lenis (weak) system of obstruents, though the actual phonetic features are not known. The strong member of each opposition was unvoiced and maybe aspirated, whereas the other member was weak and most probably also unvoiced. There are no traces of a corresponding opposition of affricates, which would mean that *č-* had no weak counterpart in Proto-Turkic.

As for word-initial stops, the Proto-Turkic system represents a secondary simplification. The earlier oppositions *k* vs. *g* and *t* vs. *d* of Proto-Turkic<sub>-1</sub> were neutralised and represented by *k-* and *t-*. As far as *p* vs. *b* is concerned, *p-* developed into a bilabial fricative and then into *h-*. Since the spirantisation of *p-* took place prior to the neutralisation of the opposition strong vs. weak with the other stops, *b-* was preserved as such. The initial *h-* disappeared in almost all languages. It remained unmarked in most Old Turkic scripts, even in cases where it still existed. Tibetan and Chinese transcriptions show that it was preserved at least in some Old Turkic dialects. The initial *h-* is also found in Khalaj (see pp. 96, 281). However, not all cases of *h-* in Old Turkic and Khalaj go back to Proto-Turkic *p-*. Some are of secondary and some are of foreign origin.

Proto-Turkic *y-* has three antecedents in an earlier Proto-Turkic<sub>-1</sub>, namely *y-*, *di-*, and *ń-*, reflected in Mongolian *y-*, *di-* and *ni-* respectively. In a large group of Turkic languages, Proto-Turkic *y-* of whatever origin changed into *j-*.

Of the two Proto-Turkic sibilants only *s-* occurred word-initially. All initial *š-* sounds in native Turkic words have developed from a primary or secondary *si-*. Proto-Turkic consonants not occurring in word-initial position are *n-*, *m-*, *ŋ-*, *l-*, *r-*, *z-*. As mentioned, however, *ń-* occurred in Proto-Turkic<sub>-1</sub>.

In non-initial positions, the consonants *p*, *b*, *k*, *g*, *t*, *d*, *č*, *s*, *š*, *m*, *n*, *ŋ*, *ń*, *r*, *z*, *l* occurred. Many old Turkic languages and dialects show spirantisation of weak obstruents, but we have no clear sources to determine when this change began.

In Chuvash and its earlier forms, *r* is found instead of *z* in other Turkic languages. This important feature of 'rhotacism' is reflected in early Turkic loanwords in Common Samoyed, Mongolian and Hungarian. While some scholars take it to reflect a special Proto-Turkic consonant or consonant combination, there is no need for such a reconstruction. Proto-Turkic had the opposition *z* vs. *r*, which was lost in the ancestor of Chuvash, but preserved

in all other Turkic languages. The loss was due to the change of *z* into *r*, a development that occurred twice in the history of Chuvash. The second time, it affected Proto-Turkic intervocalic and final *d*, which became *r* via an intermediate stage *z*. As shown by Tekin and others, *z* has developed into *r* in special, mainly pre- or post-consonantal, positions in other Turkic languages as well, e.g. *köküz* 'chest, breast' → *kökre* 'chest, upper part of the body', *semiz* 'fat' → *semri* 'become fat', *tiz* 'knee' → *tirsgek* 'elbow, knee joint' from *\*tiz-ge-k* (Hungarian *tér-d* from *\*tir* + Hungarian diminutive suffix *-d*) or Old Turkic *qoŋuz* 'beetle, insect' → *qomursga* 'ant' from *\*koŋuz* + *gA* > Chuvash *ḡāmār*. The change *z* > *r* thus began in special phonetic environments in all Turkic languages at the beginning of late Ancient Turkic. It later ceased except in the ancestor of Chuvash, where it was generalised in all positions.

In most Turkic languages, non-initial *š* corresponds to an *l* in Chuvash and its ancestors as well as in early Turkic loanwords in Common Samoyed, Mongolian and Hungarian. Scholars such as Ramstedt, Poppe and Doerfer have taken these cases of 'lambdaism' to represent a special Proto-Turkic sound or sound combination. The case is similar to that of rhotacism: Chuvash and its ancestors lost the opposition *š* vs. *l*. But not even in Chuvash was the change to *l* complete, since *š* is often found here instead of Proto-Turkic *š*, e.g. *bāš* 'head' > Chuvash *puš*. Turkic loans in Hungarian display *č* or *lč*, both going back to *č* < *š*, e.g. *bocsát* ← *\*bočat* < *bošat* 'release', *gyümölcs* ← *\*jemič* < *yemiš* 'fruit'. Similar changes are found in other Turkic languages, mainly before or after a consonant, e.g. *tapšur* 'hand over' > *tapčur*. The next stage is rare, though witnessed, e.g. Turkic *taz baši* 'sparrow hawk' > *\*tar balči* ⇒ Mongolian *tarbalji*.

### Word Structure

Word types of the structure V and CV are found in the earliest Turkic documents, e.g. *ī* 'vegetation', *tū* 'body hair'. The existence of the same structures in Proto-Turkic cannot be excluded. Often, however, it can be shown that cases of CVC had an additional vowel, which later disappeared. The basic types should have been CVCV̄ and CVCV. The reduced *-V̄* disappeared, e.g. Proto-Turkic *\*sagā* > Old Turkic *say* 'milk' ⇒ Mongolian *saya*-, Proto-Turkic *\*kertī* 'cut a notch' > Old Turkic *kert* ⇒ Mongolian *\*kerti* > *kerči*-. The non-reduced vowels remained, e.g. Proto-Turkic *teke* 'goat' > Old Turkic *teke*, Proto-Turkic *kara* 'black' > Old Turkic *qara*. The gradual loss of the final reduced vowels in Old Turkic has been demonstrated by Johanson (1979). They first disappeared after the sonorants *l*, *r* and the nasal dental *n*. Since this phase was concluded at the beginning of Old Turkic and is reflected in Chuvash, it can also be presumed for Proto-Turkic. The existence of reduced vowels after *l* is shown by some proper names. Thus, one and the same name is found in two different forms in the inscriptions: *Köl* and *Köli*.

Though the Turkic word accent is generally on the last syllable, the

disappearance of the final reduced vowels may point to a different situation in Proto-Turkic. A similar piece of evidence is the Old Turkic alliteration.

Consonant clusters were in general not possible in initial and final positions. Old Turkic final clusters such as *-rt*, *-lt*, *-lp*, *-nt*, *-nč*, *-st*, *-rk* seem to be the result of secondary developments, e.g. *and* 'oath' < Proto-Turkic *\*andā* ⇒ Mongolian *anda* 'sworn brother'.

Proto-Turkic had front vs. back sound harmony. Most features in documents and present languages contradicting this, e.g. the loss of *i* in some central Asiatic languages or the drastic changes in Chuvash, are late secondary developments and not preserved Proto-Turkic traits. While almost all suffixes had a front and a back variant, some had only one form, e.g. the front third-person singular possessive suffix *-i* and the back dative suffix *-ka*. There are also some traces of labial harmony, in a few cases even of an opposition low vs. high or low vs. reduced.

### Morphology

#### Word Formation

Proto-Turkic had various morphological devices for word formation, the most frequent formatives being suffixes. Some word formatives became unproductive very early, e.g. those in *kö-z* 'eye', *kö-r* 'see'. There was also partial reduplication as in *kap kara* 'very black' and full reduplication as in *adın adın* 'very dissimilar'.

#### Nouns

No traces of Proto-Turkic gender are found. The plural suffix was *-lar*, whereas Chuvash *-sem* is a late replacement. Proto-Turkic had several suffixes denoting collectivity, e.g. *-An* as in *oglan* 'princes, bodyguards, boys' ← *ogul* 'male child'. There was no dual. Contrary to the opinion of Erdal and others, *-z* has never been a dual suffix or denoted pairs of body parts. However, an old suffix denoting plurality can be detected in the *-z* of personal pronouns, e.g. *biz* 'we' ← *bi* 'I'.

Proto-Turkic nouns probably had an oblique stem in *-n*, just as pronouns still have in Old Turkic. Four cases show traces of it: genitive *\*-n*, accusative *\*-nVG* (in the pronominal declension *-nI*), dative *\*-nKA* and instrumental *\*-nVn*. No *-n* is found in the locative *\*-dA*. The Proto-Turkic ablative suffix *\*-dAn*, an extension of the locative, exhibits an *-n*, though not as a stem extension.

The possessive suffixes in the singular were 1p. *\*-m*, 2p. *\*-ŋ* and 3p. *\*-(s)i*. Though Chuvash has only *-i*, the existence of *-si* in the earlier history of Chuvash can be demonstrated. The plurals of the first and second persons were formed with the suffix *-z*. The third person shows no traces of a singular vs. plural opposition.