



CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

KEITH BROWN • SARAH OGILVIE



CONCISE ENCYCLOPEDIA OF
LANGUAGES
OF THE WORLD

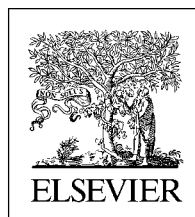
CONCISE ENCYCLOPEDIA OF LANGUAGES OF THE WORLD

COORDINATING EDITOR

KEITH BROWN
University of Cambridge
Cambridge, UK

CO-EDITOR

SARAH OGILVIE
University of Oxford
Oxford, UK



Amsterdam Boston Heidelberg London New York Oxford
Paris San Diego San Francisco Singapore Sydney Tokyo

Elsevier Ltd., The Boulevard, Langford Lane, Kidlington, Oxford, OX5 1GB, UK

© 2009 Elsevier Ltd.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic, or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publishers.

Permissions may be sought directly from Elsevier's Rights Department in Oxford, UK: phone (+44) 1865 843830; fax (+44) 1865 853333; e-mail permissions@elsevier.com.

Requests may also be completed online via the homepage (<http://www.elsevier.com/locate/permissions>).

First edition 2009

Library of Congress Control Number: 2008934269

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

ISBN 978-0-08-087774-7

09 10 11 12 13 10 9 8 7 6 5 4 3 2 1

This book is printed on acid-free paper
Printed and bound in China

Cover image: Adapted from *Orbis Terrarum Nova* (1594) by Petro Plancio

Working together to grow
libraries in developing countries

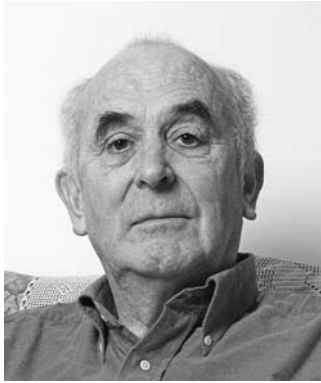
www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

THE EDITORS



Keith Brown was Editor-in-Chief of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006). He is now an Associate Lecturer in the Faculty of English at Cambridge. From 2007 he has been President of the Philological Society. From 1990 to 1994 he was President of the Linguistics Association of Great Britain, and he has been a Member of Council of the Philological Society since 1998. He is author of *Linguistics Today* (Fontana, 1984) and co-author, with Jim Miller, of *Syntax: A Linguistic Introduction to Sentence Structure* and *Syntax: Generative Grammar* (Hutchinson, 1981). Keith was joint editor of *Concise Encyclopedia of Linguistic Theories* and *Concise Encyclopedia of Grammatical Categories* (Pergamon Press, 1997 and 1998), *Common Denominators in Art and Science* (Aberdeen University Press, 1983) and *Language, Reasoning and Inference* (Academic Press, 1986).



Sarah Ogilvie, Trinity College, Oxford, is a linguist and lexicographer who specializes in words that enter English from non-European languages. She was Languages of the World section editor of the second edition of the *Encyclopedia of Language and Linguistics* (Elsevier, 2006), a former editor of the *Oxford English Dictionary*, and was Etymologies Editor of the *Shorter Oxford Dictionary* (6th ed., 2007).

- Rodrigues A D (2000). 'Gê-Pano-Carib × Jê-Tupí-Karib: sobre relaciones lingüísticas prehistóricas en Sudamérica.' In *Actas del I Congreso de lenguas Indígenas de Sudamérica*. Lima: Universidad Ricardo Palma 1. 95–104.
- Rodrigues A D & Cabral A S A C (2002). 'Reverendo a classificação interna da família Tupi-Guaraní.' In *Atas do I Encontro Internacional do Grupo de Trabalho de Línguas Indígenas da Associação*

Nacional de Pós-Graduação em Letras e Linguística (ANPOLL), tomo I. Belém: Editora Universitária UFPA. 327–337.

- Rodrigues A D & Dietrich W (1997). 'On the linguistic relationship between Mawé and Tupi-Guarani.' *Diachronica* 14, 265–304.
- Storto L R (1999). Aspects of a Karitiana grammar. Ph.D. diss., MIT.

Turkic Languages

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Development and Classification

The Turkic language family was first attested in 8th century inscriptions. Turkic-speaking groups first appeared in the Inner Eurasian steppes, from where they moved to Central Asia, Eastern Europe, the Middle East, Siberia, etc. Because of their high mobility, Turkic expanded over a huge area.

The Proto-Turkic network of varieties was dissolved by an early split of Oghur or Bulgar Turkic. Its modern representative, Chuvash, a descendant of Volga Bulgar, differs from Common Turkic by specific phonetic representations, e.g., *r* and *l* instead of *z* and *š* in words such as *šér* 'hundred' and *šul* 'year' (Turkish *yüz* 'hundred,' *yaş* 'age'). A second split is represented by Khalaj, which retains a reflex of Proto-Turkic **p*- as *h*-, e.g., *badaq* 'foot.' Dialect splitting has led to further differentiation of Common Turkic. There is no mutual intelligibility throughout the family today. The following division combines the current areal distribution with genealogical and typological features.

1. The Southwestern or Oghuz branch contains a western subgroup comprising Turkish, Gagauz, and Azerbaijani (Azerbaijani, Northern and Azerbaijani, Southern), a southern subgroup comprising dialects of southern Iran and Afghanistan, and an eastern subgroup comprising Turkmen and Khorasan Turkic.
2. The Northwestern or Kipchak branch has a western subgroup comprising Kumyk, Karachay-Balkar, Crimean Tatar, and Karaim, a northern subgroup comprising Tatar and Bashkir, and a southern subgroup comprising Kazakh, Karakalpak, Kipchak Uzbek, Nogai, and Kirghiz (of different origin, but strongly influenced by Kazakh).

3. The Southeastern or Uyghur-Karluk branch has a western Uzbek subgroup and an eastern Uyghur subgroup.
4. The Northeastern or Siberian branch has a southern heterogeneous subgroup comprising Sayan Turkic (Tuvan, Tofan), Abakan (Yenisei) Turkic (Khakas, Shor), Chulym Turkic, Altai Turkic (Altai, Northern and Southern), and a northern subgroup comprising Yakut (Sakha) and Dolgan.
5. Chuvash is geographically situated in the northwestern area (Volga region).
6. Khalaj is geographically situated in the southwestern area (central Iran).

Deviant languages in China are Salar, of Oghuz origin, Yellow Uyghur (Yugur, West) and Fu-yü (Manchuria), both of south Siberian origin.

One traditional classificatory criterion is the final consonant of the word for 'nine.' Its representation as *r* in Chuvash *täxxär* separates Oghur from Common Turkic (Turkish *dokuz*). The intervocalic consonant in the word for 'foot' divides most Northeastern languages, Chuvash, Khalaj, etc. from the rest, which exhibits *-y-* (Turkish *ayak*), e.g., Tuvan *adaq*, Khakas *azax*, Chuvash *ura*. Oghuz Turkic differs from the rest by loss of suffix-initial velars, e.g., *qal-an* [remain-PART] instead of *qal-yan* [remain-PART] 'remaining.' Final *-G* is devoiced in the Southeast (Uyghur *tay-liq* [mountain-DER] 'mountainous'), preserved in southern Siberia (Tuvan *day-lïy* [mountain-DER]), and lost elsewhere (Turkish *dağ-lı* [mountain-DER]).

Most older linguistic stages are insufficiently known. Written sources, where available, provide no direct information on spoken varieties. Early Oghuz and Bulgar (East Europe, 6th–7th centuries) are unknown. There are no texts in the language of the Khazars (7th–10th centuries). Pecheneg and Kuman, predecessors of West Kipchak, are only known from loanwords, titles, and names.

Written Varieties

Turkic literary varieties have emerged in various cultural centers. Many older Turkic empires, however, used foreign languages for administration (Sogdian, Persian). Muslim Turks often used Persian for poetry, and Arabic for religious and scientific writing. Russian has played an important role for many groups. The following main stages of written Turkic may be distinguished.

1. An older pre-Islamic East Old Turkic period (8th century–), is represented in inscriptions, manuscripts, and block prints. East Old Turkic proper is documented in stone inscriptions (Orkhon Valley), which celebrate the rulers of the Second Eastern Türk Empire, in other inscriptions found in Mongolia and the Yenisei and Talas valleys, and also in a few manuscripts. The Old Kirghiz inscriptions are of this type. Old Uyghur is first recorded in the period of Uyghur rule over the Eastern Empire. Early Old Uyghur is attested in runiform inscriptions and manuscripts. From the 10th century on, Old Uyghur became the medium of a flourishing literary culture in the Tianshan-Tarim area, attested in texts of Buddhist, Manichaean, and Nestorian content.
2. A middle Turkic period comprises various early Islamic varieties.

The first East Turkic written language, Karakhanid (11th century–), developed in Kashgar, is close to Old Uyghur but lexically influenced by Arabic and Persian. Maḥmūd of Kashgar provides information (1073) on Karakhanid and other contemporary Turkic varieties.

Khorezmian Turkic, used in the 13th–14th centuries in the Golden Horde and Mamluk Egypt, is based on the older languages but contains Oghuz and Kipchak elements.

This tradition is continued in Chaghatay (15th century–). Early Chaghatay contains regional elements of the Timurid area. Later, Chaghatay became the dominant written language of Central Asia, eventually conquering an immense area of validity and developing regional varieties.

The first West Turkic written language is Volga Bulgar, insufficiently known from epitaphs of the 13th and 14th centuries. Information on early Kipchak Turkic is given in the *Codex Cumanicus*, compiled by Christians, and in dictionaries and grammars written in Mamluk Egypt and Syria.

Oghuz Turkic is first represented by Old Anatolian Turkish (13th century–), which was a subordinate written medium until the end of Seljuk rule. Old Ottoman is the initial stage of Ottoman, which

begins with the foundation of the Ottoman Empire in 1307. In Azerbaijan a literary language developed from the 15th century on.

3. A premodern period (16th century–) begins with the development of regionally influenced written languages. Middle and Late Ottoman became the leading written language with an abundantly rich literature. Chaghatay continued to play a major role and remained the literary language of all non-Oghuz Muslim Turks until a century ago.
4. A modern period begins in the second half of the 19th century with the formation of regional written languages. The political division of the Turkic-speaking world in the 20th century and the language policies pursued in the Soviet Union, Turkey, China, and Iran had dramatic effects that increasingly obstructed transregional linguistic contacts. A dozen ‘national’ languages with a narrow radius of validity emerged. In Turkey, Ottoman was replaced by modern Turkish. The social importance of many Turkic languages was very limited. After the recent political developments, their significance is rapidly increasing, but the varieties spoken in Iran, Afghanistan, Iraq, etc., still have poor possibilities to develop.

Various scripts and script systems have been applied to Turkic. A specific runiform script was created for Old East Turkic. Most Old Uyghur texts are written in Uyghur script, originating in the Near East and later taken over by Mongols and Manchus. It is similar to the Sogdian script, which is also used in Buddhist texts. A few Buddhist manuscripts are written in Brahmi script, Manichaean texts in Manichaean script, and Nestorian texts in Syriac script. Arabic script was used for the languages of the Islamic era (still used in China for Uyghur and Kazakh). A unified Roman-based script was introduced for several languages in the early Soviet period, but later replaced by different Cyrillic-based scripts. A Roman-based alphabet was introduced in Turkey in 1923. Most of the newly established Turkic republics have introduced or are introducing Roman-based scripts.

Contacts

The massive displacements of Turkic-speaking groups throughout their history have led to various phenomena induced by contacts with Iranian, Slavic, Mongolic, Uralic, etc. Speakers of Turkic have copied lexical, phonetic, morphological, and syntactic elements, whereas non-Turkic (e.g., Iranian, Greek, Finno-Ugric, Samoyedic, Yeniseian, Tungusic) groups shifting to Turkic have exerted substrate influence by

copying native elements into their new varieties. Languages such as Chuvash, Yakut, Salar, Yellow Uyghur, Khalaj, Karaim, and Fu-yü have long developed in isolation from their relatives, preserving old features and acquiring new ones in their environments. Long and intense interaction with Iranian in Central Asia, Iran, Afghanistan, etc., has led to profound convergence phenomena. Massive foreign influence has sometimes caused considerable typological deviations, e.g., drastic structural changes in Karaim and Gagauz under Slavic impact.

Most written languages have been strongly influenced by Persian and Arabic. In Chaghatay (Chagatai) and Ottoman, lexical borrowing contributed to a remarkable richness of the vocabularies, whereas grammar was much less affected. The overload of Persian and Arabic in Ottoman led to strong puristic efforts in the 20th century to create a so-called Pure Turkish.

Internal convergence processes have resulted in leveling of languages of the central area. Several Turkic koinés have been used as transregional codes for trade and intergroup communication, e.g., Azerbaijani in Iran and the Caucasus region.

Linguistic Features

Despite their huge area of distribution, Turkic languages share essential phonological, morphological, and syntactic features.

They have a synthetic word structure with numerous highly applicable derivational and grammatical suffixes, and a juxtaposing technique with clear-cut morpheme boundaries and predictable allomorphs. These agglutinative principles yield considerable morphological regularity and transparency. Exceptions include traces of vowel gradation in the pronominal declination, e.g., Turkish *ben* 'I,' *ban-a* [I-DAT] 'to me.' The agglutinative structure is partly deranged in languages of the northeast and southeast. Some languages, e.g., Uzbek, even display borrowed prefixes.

The syllable contains minimally a vowel with maximally one preceding and one subsequent consonant. Vowel hiatus and consonant clusters are avoided.

Most languages exhibit eight short vowel phonemes, *a, i, o, u, e, i, ö, ü*, classified according to the features front vs. back, unrounded vs. rounded, and high vs. low. Proto-Turkic long vowel phonemes are preserved in Turkmen, Yakut, and Khalaj. Iranian and Slavic phonetic influence has sometimes affected the front vs. back distinctions. Tatar, Bashkir, Chuvash, and Uyghur exhibit systematic vowel shifts. Chuvash, Gagauz, Karaim, etc., have developed palatalized

consonants, e.g., Karaim *meni* 'I'. Tuvan and Tofan exhibit a glottal element signaling strong obstruents, e.g., *aʔt* 'horse' vs. *at* 'name.'

The most general sound harmony phenomenon is an intrasyllabic front vs. back assimilation. An inter-syllabic front vs. back harmony causes neutralization of the front vs. back distinction under the influence of the preceding syllable. If applied consistently, it excludes back and front syllables in a word, e.g., Turkish *ev-ler-im-e* [house-PL-POSS.1.SG-DAT] 'to my houses,' *at-lar-ım-a* [horse-PL-POSS.1.SG-DAT] 'to my horses.' Some languages only display this kind of harmony, whereas others also apply a rounded vs. unrounded harmony, neutralization of the distinction rounded vs. unrounded in high suffix vowels, e.g., Turkish *el-im* [hand-POSS.1.SG] 'my hand,' *gül-üm* [rose-POSS.1.SG] 'my rose.' Languages such as Yakut and Kirghiz apply this harmony to low-vowel suffixes as well, e.g., *börö-lör* [wolf-PL] 'wolves.' There are numerous exceptions to harmony rules in loanwords. Further allomorphs are created by various consonant assimilations.

The rules of word accent vary. A high pitch accent, interacting with a dynamic stress accent, mostly falls on the last accentable syllable of native words.

The morphological structure has remained relatively stable through the centuries. The main word classes are nominals (nouns, adjectives, pronouns, numerals) and verbals. The primary stems can be used as free forms, e.g., *at* 'horse,' *at!* 'throw!'. From verbal and nominal stems, which are sharply distinguished, expanded stems are formed. Nominals take plural, possessive, case, and specific derivational suffixes. Grammatical gender is not marked. The verbal morphology comprises markers of actionality, voice, possibility, negation, aspect, mood, evidentiality, tense, person, interrogation, etc. Voice is expressed by passive, reflexive-middle, causative, and cooperative-reciprocal suffixes. The order and combinability of suffixes is basically common to all Turkic languages.

Constructions with postposed auxiliary verbs (post-verbs) express actional modifications. A few constructions have developed into aspect-tense categories, e.g., Turkish *gel-iyor* [come-PRES] 'comes' < **gel-e yori-r* [come-CONV run-AOR] ('runs coming'). Possibility markers are formed with auxiliary verbs such as *bil-* 'to know' and *al-* 'to take,' e.g., Kirghiz *ber-e al-* [give-CONVAUX.POTEN] 'to be able to give.'

Turkic languages share many syntactic characteristics. With respect to relational typology, they adhere to the nominative-accusative pattern. They have a head-final constituent order, with dependents preceding their heads. The unmarked order of clause constituents is subject + object + predicate (SOV).

Adjectival, genitival, and participial attributes precede the head of the nominal phrase. Postpositions are used instead of prepositions. There is no agreement in number or case between dependents and heads. The focus position is in front of the predicate core. The unmarked constituent order is often deviated from for discourse-pragmatic reasons. Contact-induced word order changes are common, e.g., in Gagauz, which has become an SVO language.

Preposed subordinate clauses are based on verbal nouns, participles, and converbs. The use of postposed subordinative patterns with conjunctions are typical effects of Iranian and Slavic influence. Most languages possess conjunctions, even coordinative ones meaning 'and,' 'or,' and 'but' of Persian, Arabic, or Russian origin.

Turkic lacks definite articles. The indefinite article is formally identical with the numeral 'one' Genitival attributes, expressing a possessor, stand in the genitive, whereas their head, indicating a possessed entity, carries a possessive suffix, e.g., Turkish *at-in baş-ı* [horse-GEN head-POSS.3.SG] 'the head of the horse.' The dominant type of nominal compounds follows the

pattern noun + noun + possessive suffix, e.g., Turkish *el çanta-sı* [hand bag-POSS.3.SG] 'handbag.'

All Turkic varieties exhibit numerous loanwords. Arabic and Persian loans are frequent in all Islamic-Turkic languages. The Iranian influence is strong in Uyghur, Uzbek, and varieties of Iran and Afghanistan. Many languages have been subject to considerable Mongolic and Slavic influence. Loans and calques from European languages have become increasingly important. The Turkic languages spoken in China exhibit old and recent loans from Chinese.

Bibliography

- Deny J *et al.* (eds.) (1959). *Philologiae turcicae fundamenta*, 1. Aquis Mattiacis: Steiner.
- Johanson L (2002). *Structural factors in Turkic language contacts*. London: Curzon.
- Johanson L & Csató É Á (eds.) (1998). *The Turkic languages*. London: Routledge.
- Menges K H (1995). *The Turkic languages and peoples. An introduction to Turkic studies*. Wiesbaden: Harrassowitz.
- Róna-Tas A (1991). *An introduction to Turkology*. Szeged: University of Szeged.

Turkish

R Underhill, San Diego State University, San Diego, CA, USA

© 2006 Elsevier Ltd. All rights reserved.

Turkish (natively *Türkçe*), the official language of the Republic of Turkey, is spoken by a large proportion of the Turkish population. There are also Turkish speakers in the Balkans, particularly in Greece, Bulgaria, and the former Yugoslavia, although there has been extensive population inflow from those countries into Turkey, and there is a substantial minority of Turkish speakers in Cyprus. There are Turkish-influenced Turkic dialects in Iraq in the region of Kirkuk, where the speakers are called Turkmen or Turkomans. The *Ethnologue* entry for Turkish gives a population of roughly 46 million speakers in Turkey, and 61 million in all countries.

Turkish belongs to the southwestern, or Oghuz (*Oğuz*), group of Turkic languages. This group also includes Azerbaijani, spoken in Azerbaijan and in adjacent areas of Iran; Qashqay and related dialects, spoken in the Zagros mountain area of Iran; Türkmen, spoken in Turkmenistan; and Gagauz,

spoken in Bulgaria, in Romania, and principally in Moldova, although there has been substantial migration from Moldova to Turkey. Central Asian Turkic languages include the national languages of Kazakhstan, Uzbekistan, and Kyrgyzstan, and a number of others. Turkic, in turn, belongs to the Altaic family of languages, which also includes the Mongol and Manchu-Tunguz language families. Though this relationship has recently been called into question, it was proved convincingly by Poppe more than a generation ago (Poppe, 1960). Wider affinities of the Altaic family have been suggested for Korean, and even for Japanese.

Turkish scholars divide the history of the Turkish language into three periods: (1) Old Anatolian Turkish (*Eski Anadolu Türkçesi*), comprising texts dating from the earliest arrival of Turkic speakers in Anatolia, through the Seljuk period to the formation of the Ottoman Empire; (2) Ottoman (*Osmanlıca*), the language of the Ottoman Empire, heavily influenced by Arabic and Persian; and (3) Modern Turkish (*Yeni Türkçe*), dating from the overthrow of the Ottoman Empire and from the Turkish language reform movement of the 1920s and 1930s. The Turkish language reform movement was

launched by Atatürk as part of his overall plan to distance Turkey from Middle Eastern, specifically Arabic and Persian, influences, in favor of European influence. This movement in the language area included most noticeably the replacement of the Arabic writing system with a Latin alphabet in 1928, and a drive to replace Arabic and Persian vocabulary, once pervasive in Ottoman texts, with vocabulary drawn or constructed from Turkish sources, or Turkish-looking inventions. The drive to cleanse the lexicon has waxed and waned over the interim and has acquired political correlates: writers on the left tend to use neologisms; those on the right use a more traditional vocabulary. There has been no corresponding attempt to rid the lexicon of European or English terminology (for more on the language reform movement, see Lewis (1999)). In 1997, a committee of the American Association of Teachers of Turkic Languages attempted to create a standardized English terminology for Turkish, which is used here.

Phonology

Phonemes

Consonants The International Phonetic Association (IPA) representations of the Turkish consonant system are shown in Table 1. Turkish uses 21 letters for consonants: b c ç d f g ğ h j k l m n p r s ş t v y z. These represent the expected sounds, except as follows:

Letter	Sound
c	[dʒ]
ç	[tʃ]
j	[ʒ]
ş	[ʃ]

In the following discussions, [tʃ] and [dʒ] will henceforth be written /ç/ and /j/, since they function in all phonological respects as members of the natural class of stops, not as clusters. The letters *k g l* each stand for two sounds: a plain velar or lateral [k g l] and a front velar or palatal [c ʝ λ]. In words of Turkish origin, the front velar variant occurs with front vowels and the plain velar occurs with back vowels.

Table 1 International Phonetic Association symbols for Turkish consonants

Labial	Dental	Palatal	Front velar	Velar	Glottal
p	t	tʃ	c	k	
b	d	dʒ	ʝ	g	
f	s	ʃ			h
v	z	ʒ			
m	n	ɲ			
	r	ʀ			

In words of Arabic origin, however, /c ʝ λ/ can occur with back vowels, giving rise to pairs and thus distinctive contrasts, as in *kar* ‘snow’ [kar] and *kâr* ‘profit’ [car].

The letter ğ, or *yumuşak ge* ‘soft g’, has no consonantal sound. It normally represents an historical or underlying /g/ that has been deleted; in some Anatolian dialects, it survives as a voiced fricative [ɣ]. Most commonly, ğ lengthens the preceding vowel in syllable-final (coda) position, and represents nothing between vowels, as in *dağ* ‘mountain’ [da:] and *dağa* ‘mountain (dat)’ [daa].

Vowels Turkish vowels are traditionally represented in a ‘cube’ shape, consisting of all possible values of the features, front/back, high/low, and rounded/unrounded, as in Figure 1. Each vowel can occur long, from the deletion of ğ, and the vowels /e i a u/ can occur long in Arabic loanwords, giving a total of 16 vowel phonemes. The vowel letters are for the most part self-explanatory, except for *ı*, an undotted ‘i,’ which is a high back unrounded vowel, IPA [u]. All Turkish vowels are phonetically lax, except sometimes before *y* or *ğ*, thus *a e i i o ö u ü* sound like [a ε ɪ u u ɔ œ ʊ ʏ]. Because the difference between *ı* and *i* is distinctive, it must be maintained for capitals also, i.e., *I* and *İ*.

Stress Stress in Turkish consists of higher pitch, rather than greater loudness on the accented syllable. Stress is normally on the last syllable of the word; as affixes are added, stress moves rightward:

- (1) *él* ‘hand’
- ellér* ‘hands’
- ellerím* ‘my hands’

There are a number of exceptions to final stress. Some words have inherent nonfinal stress, and in these cases stress does not move with the addition of affixes. Inherently stressed words include most loans, which have their own rule for accent; in such cases, the accent may fall on a syllable other

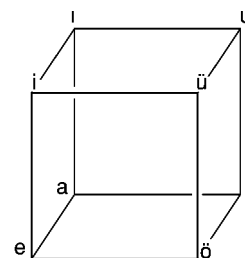


Figure 1 Turkish vowels. Front vowels are represented at the front of the cube, high vowels are at the top, and rounded vowels are to the right. Reproduced from Underhill R (1976) Turkish grammar. Cambridge: MIT Press. With kind permission by MIT Press.

than that which is stressed in the source language, as in *sinéma* ‘cinema’ and *Kenédi* ‘Kennedy’. Some affixes are prestressing; stress then falls on the preceding syllable and remains there as additional affixes are added. The rules for stress and much else in Turkish phonology are extensively worked out in Demircan (2001).

Phonological Rules

Turkish being an agglutinating language, suffixes are added to stems in such a manner that segmentation is relatively easy. However, a number of changes take place in both stems and suffixes when this happens.

Vowel Harmony Vowel harmony involves the two features front/back and rounded/unrounded. It is a syllable-to-syllable process by which each vowel conditions the following vowel, according to the following rules:

1. Any of the vowels can occur in the first syllable of a word.
2. A noninitial vowel assimilates to the previous vowel in frontness.
3. A noninitial high vowel assimilates to the previous vowel in rounding. A noninitial low vowel is unrounded. Thus /o ö/ do not appear in harmonic suffixes.

The process is illustrated in Table 2, which shows how the stem, dative (suffix *-yA*), and objective (suffix *-yI*) case forms of a set of nouns are used (in morphophonemic transcription, the symbol *A* represents the alternation between /a/ and /e/, and the symbol *I* represents the alternation /i ı u ü/). A few native words and very many foreign words are nonharmonic, such as *kardeş* ‘brother’, *otel* ‘hotel’, and *sigorta* ‘insurance’. This has led some scholars to claim that vowel harmony no longer holds for stems (Clements and Sezer, 1982). In the case of a nonharmonic word, suffixes are controlled by the last syllable, as in *asansör* ‘elevator’ (plural *asansörler*) and *kredikart* ‘credit card’ (plural *kredikartlar*).

Table 2 Turkish vowel harmony

Stem	Gloss	Dative	Objective
<i>bal</i>	‘honey’	<i>bala</i>	<i>balı</i>
<i>kıl</i>	‘hair’	<i>kıla</i>	<i>kılı</i>
<i>ok</i>	‘arrow’	<i>oka</i>	<i>oku</i>
<i>buz</i>	‘ice’	<i>buzı</i>	<i>buzu</i>
<i>ev</i>	‘house’	<i>eve</i>	<i>evı</i>
<i>il</i>	‘province’	<i>ile</i>	<i>ilı</i>
<i>göl</i>	‘lake’	<i>göle</i>	<i>gölü</i>
<i>gül</i>	‘rose’	<i>güle</i>	<i>gülü</i>

Other Phonological Rules Beyond vowel harmony, stems and suffixes have a highly changeable nature. Suffix-initial voiced stops devoice after a stem ending in an unvoiced consonant. Many suffixes have different postconsonantal and postvocalic forms. Stems also undergo a number of rules designed to maintain canonical syllable structure, particularly in closed syllables. Among the rules applying to syllables are final devoicing, epenthesis, degemination, and vowel shortening. There are many details concerning these rules, but as an extreme example, the verbal noun suffix best written as *-DIg* has 16 forms:

-dik/dık/duk/dük/tik/tık/tuk/tük/diğ/dığ/duğ/düğ/tiğ/tığ/tuğ/tüğ

Morphology

Turkish is an agglutinating language in which suffixes, in some cases a large number of them (the lists of suffixes in the following sections are not exhaustive), are added fairly transparently to stems:

- (2) *ev* ‘house’
evler ‘houses’
evlerim ‘my houses’
evlerimiz ‘our houses’
evlerimizde ‘in our houses’
evlerimizdeki ‘which is in our houses’

The Noun Paradigm

Noun stems may have the following inflectional suffixes, in order:

1. Plural *-IAr* (as in *baba* ‘father’, *babalar* ‘fathers’ and *deve* ‘camel’, *develer* ‘camels’).
2. Possessive (possessed agreement).
3. Case (as in *oda* ‘room’).

- (3) Nominative: *oda*
 Genitive (-*(n)In*): *odanın*
 Dative (-*yA*): *odaya*
 Objective (-*yI*): *odayı*
 Locative (-*DA*): *odada*
 Ablative (-*DAn*): *odadan*
 Instrumental/comitative (-*y-IA*): *odayla*

The Verb Paradigm

Starting with the verb root, a number of derivational suffixes can be added to build up the verb stem. These include reflexive, reciprocal, causative, passive, impossibility, negative, and abilitative forms. At this point, from the verb stem, it is possible to go in a number of directions. For a finite (‘tensed’) verb, the next step is a tense suffix, followed normally by a personal ending:

(4) General present:	gelirim	'I come', 'I'll come'
Progressive:	geliyorum	'I am coming'
(Definite) past:	geldim	'I came'
Unwitnessed past:	gelmişim	'I (supposedly) came'
Future:	geleceğim	'I will come'
Necessitative:	gelmeliyim	'I ought to come'
Optative:	geleyim	'let me come'
Conditional:	gelsem	'if I come'

There is also a wide range of nonfinite suffixes possible at this point for the formation of subordinate clauses. These include verbal nouns or nominalizations, participles, and adverbial clause suffixes (traditional 'converbs').

Auxiliary Suffixes

Finally, there is a group of suffixes that can be categorized under the heading of 'auxiliary'. They can be added both to verbal and nonverbal predicates, hence a separate auxiliary category. They include most prominently the personal endings, but also some morphemes that can be called 'aspects', although they are not all aspects any more than the tenses are all tenses (abbreviations: SG, singular; PROG, progressive):

- (5) Yorgun -du -m.
tired -PAST -1SG
'I was tired'.
- (6) Gel -iyor -du -m.
come -PROG -PAST -1SG
'I was coming'.

The aspects are past *-y-DI*, dubitative *-y-mIş*, and conditional *-y-sA*. Furthermore, there is an adverbial aspect *-y-ken*. These look very similar to some tenses, i.e., definite past *-DI*, unwitnessed past *-mIş*, and conditional *-sA*, but they differ in morphology, meaning, and prosody (all auxiliary suffixes are prestressing).

The inferential/quotative, sometimes called dubitative (DUB), *-y-mIş*, deserves special discussion. This aspect, and to some extent the corresponding tense, *-mIş*, are used when the speaker wishes to be disassociated from the truth of the utterance – for example, when the speaker has information that has only been heard or recently found out (VB, verb):

- (7) Sen tembel -miş -sin.
you lazy -DUB -2SG
'They say you are lazy'.
- (8) Geçen sene hasta -lan-mış-sın.
past year sick -VB-DUB-2SG
'(I heard) you got sick last year'.

The dubitative can also be used for statements for which the speaker does have personal knowledge of

the fact, but is expressing something unexpected or surprising – for example, after trying a food that the speaker had expected to dislike:

- (9) Bu yemek iyi -miş!
this food good -DUB
'This food is good!'

Syntax

Unmarked (normal) word order is subject-object-verb, as shown in the following example (OBJ, objective; DAT, dative):

- (10) Hasan mektub -u
Hasan letter -OBJ
Ayşe-ye gönder -di.
Ayşe-DAT send -PAST
'Hasan sent the letter to Ayşe'.

However, this is complicated by the fact that Turkish has pragmatically conditioned word order, by which the information status of noun phrases, rather than their grammatical function, determines their placement in the sentence. Many of the basic principles were worked out by Erguvanlı (1984). The topic is sentence initial; thus, any of the terms of Example (10) could be initial, depending on whether Hasan, the letter, or Ayşe is the topic. New information comes in the preverbal position, thus any of the terms of Example (10), if indefinite, would move preverbally:

- (11) Mektub-u Ayşe-ye bir
letter-OBJ Ayşe-DAT a
arkadaş gönder-di.
friend send-PAST
'A friend sent the letter to Ayşe'.

In fact, preverbal position is focus position; thus, *wh*-words are found here, as well as words questioned contrastively, the focused words in the answers to *wh*-questions, or any focused argument. Though the canonical sentence pattern for English might be written as subject-verb-object-X, where X is everything else, the pattern for Turkish would be topic-X-focus verb, and is thus determined by pragmatic rather than by grammatical conditions. Furthermore, sentences are not necessarily verb final. Backgrounded or unstressed information can move to the right of the verb, producing what is traditionally called a *devrik cümle (tüümce)*, or 'inverted sentence' (NEG, negative; PL, plural):

- (12) Ver-me çocuğ-a kibrit-ler-i.
give-NEG child-DAT match-PL-OBJ
'Don't give the child the matches'.

The focus in Example (12) is 'don't give,' and the child and the matches will have been previously

mentioned or are clear in the context, i.e., are ‘given’ in the sense of functional syntax.

Turkish is a left-branching and head-final language in which nouns follow adjectives (Example (13)), possessives (Example (14)), and relative clauses (Example (15)); postpositions follow noun phrases (Example (16)), and verbs follow direct objects, even subordinate clauses (Example (17)) (GEN, genitive; POSS, possessive; LOC, locative; PART, participle; ABL, ablative; VN, verbal noun; FUT, future):

- (13) çok küçük bir çocuk.
very small a child
‘A very small child’.
- (14) Enver-in şapka-sı.
Enver-GEN hat-POSS
‘Enver’s hat’.
- (15) Köşe-de otur-an kız.
corner-LOC sit-PART girl
‘The girl who is sitting in the corner’.
- (16) Bu haber-den dolayı.
this news-ABL because
‘Because of this news’.
- (17) Hasan-in yarın
Hasan-GEN tomorrow
gel-eceğ-in-i duy-du-m.
come-VN.FUT-3SG-OBJ hear-PAST-1SG
‘I heard that Hasan will come tomorrow’.

Notice from Example (17) that Turkish is a pro-drop language (‘pronoun dropping’; i.e., subject pronouns

normally are not used, as in Latin or Spanish). Overt pronouns appear in cases of focus or contrast, including topic change. Because relative clauses precede head nouns, and direct objects (including noun complement clauses) precede the main verb, Turkish sentences sometimes give the impression of having the reverse word order from English. English speakers reading Turkish sometimes find it easier to start at the end of a sentence and read toward the front, and Turkish speakers report that they do the same in reading English.

Bibliography

- Clements G N & Sezer E (1982). ‘Vowel and consonant disharmony in Turkish.’ In van der Hulst H & Smith N (eds.) *The structure of phonological representations, part II*. Dordrecht: Foris. 213–255.
- Demircan Ö (2001). *Türkçenin ses dizimi*. İstanbul: Der Yayınları.
- Erguvanlı (Taylan) E (1984). *The function of word order in Turkish grammar*. Berkeley: University of California Press.
- Kornfilt J (1997). *Turkish grammar*. London: Routledge.
- Lewis G (1999). *The Turkish language reform: a catastrophic success*. Oxford: Oxford University Press.
- Poppe N (1960). *Vergleichende grammatik der Altaischen sprachen*. Wiesbaden: Harrassowitz.
- Taylan E E (2001). *The verb in Turkish*. Amsterdam: John Benjamins.
- Underhill R (1976). *Turkish grammar*. Cambridge: MIT Press.

Turkmen

L Johanson, Johannes Gutenberg University, Mainz, Germany

© 2006 Elsevier Ltd. All rights reserved.

Location and Speakers

Turkmen (*türkmen dili*, *türkmençe*) belongs to the southwestern or Oghuz branch of the Turkic language family, which also includes Turkish. It is mainly spoken in Turkmenistan (*Türkmenistan döwleti*), which is located in the Transcaspien region and whose capital is Ashgabat. Turkmenistan borders on Iran and Afghanistan in the south, Uzbekistan in the east, and Kazakhstan in the north. The area of distribution of Turkmen extends from the southeastern shore of the Caspian Sea to the Kazakh-speaking area in the north, the Karakalpak-speaking area in the north-east, the Uzbek-speaking area in the east, beyond

the Amudarya River, and the Persian (Farsi, Western) and Khorasan Oghuz (Khorasani Turkish) areas in the south, beyond the borders to Afghanistan and Iran. Though Turkmen make up 85% of the 4.8 million inhabitants, only 72% speak Turkmen. The other main languages are Russian (12%) and Uzbek (9%). Turkmen-speaking groups also live in the Russian Federation, Kazakhstan, Tajikistan, China, etc. The total number of speakers amounts to nearly 5 million.

The designation ‘Turkmen’ is not unequivocal. Older Oghuz varieties spoken in Khorezm, Khorasan, Azerbaijan, Anatolia, and other regions in the Near East were referred to as ‘Turkmen.’ Several nomadic groups in Anatolia, Iraq, etc. are still called ‘Turkmen’ without being Turkmen in a linguistic sense.

Since the mid-1990s, language policy aims at consolidating Turkmen as the state language and to remove the Russian dominance. Turkmen is gaining

more social functions. The 1992 constitution defines it as the “official language of inter-ethnic communication.” Geographic names and administrative terms have been changed from Russian to Turkmen. In practice, however, Russian has maintained its importance in most spheres of public communication.

Origin and History

The Turkmen go back to the Turkic-speaking Oghuz confederation of tribes, whose Inner Asian steppe empire collapsed in 744. Certain Oghuz groups migrated into the region between the Syrdarya and Ural rivers. By the late 10th century, the Seljuk dynasty was founded, and an autonomous state was established on the lower Syrdarya. The Seljuks left this region in the middle of the 11th century and migrated westwards. Their modern descendants are the Turks of Khorasan, Azerbaijan, and Turkey. The speakers of Turkmen are mainly descendants of non-Seljuk groups that did not take part in these migrations.

During the Mongol conquests in the 13th century, the remaining Oghuz tribes were pushed into the Karakum desert and the region east of the Caspian Sea. From the 16th century on, Turkmen groups migrated to Khorezm, to the southern part of today’s Turkmenistan, and to Khorasan, absorbing local Turkic and Iranian elements. The major migrations of the Salīr, Ersari, Sariq and Teke tribes took place in the 17th century. In the 18th century, the Turkmen conquered the whole core area that they inhabit today.

Most tribes were subsequently divided and controlled by the Uzbek khanates of Khiwa and Bukhara, while the Persian shahs tried to subdue the southern tribes. The dependence of Khiwa and Persia came to an end after the mid-19th century. Some decades later, Russia annexed the Turkmen territory, which caused many Turkmen groups to emigrate to Afghanistan and Iran. The Turkmen area was first administered as the Trans-Caspian district in the Governorate of Turkistan. In 1924, Turkmenistan was proclaimed a Socialist Soviet Republic.

In connection with the dissolution of the Soviet Union, Turkmenistan declared its sovereignty in 1990, achieved its independence in 1991 (after a popular referendum), and adopted its new constitution in 1992.

Related Languages and Language Contacts

The closest relative of Turkmen is Khorasan Turkic (Khorasani), spoken in northeastern Iran and Khorezm, a distinct language with which it constitutes the eastern subbranch of Oghuz. Azerbaijanian

(Azerbaijani) and Turkish represent the western subbranch. The specific features of Turkmen are partly archaic and partly innovative, due to language contact. Within the Turkic family, Khorasan Turkic, Uzbek, and Karakalpak are the most important contact languages. Turkmen has had intensive contacts with Persian and, during the last century, with Russian.

The Written Language

Old Turkmen is not clearly documented in written sources. The oldest records of ‘Turkmen’ relate to Oghuz varieties in general. Oghuz texts of the following centuries do not exhibit any specific Turkmen features. A written Turkmen literature began in the 18th century, but the language used is a variety of the classical Chaghatay (Chagatai) language. A Turkmen standard language was created in the Soviet era and formed mainly from 1928 on. It was based on the Teke dialect as spoken in the Ashgabat region.

Arabic script was used in the first period. Two script reforms, in 1922 and 1925, aimed at reflecting spoken features more adequately. A Roman-based alphabet that reflected most of these features rather accurately was in use from 1928 to 1940. A variant of the Cyrillic alphabet was adopted in 1939–1940. Since the early 1990s, there has been a transition to a Roman-based script again. In 1993, the final version of a Roman-based alphabet was adopted to replace the Cyrillic one. It has several unique letters that distinguish it from Turkey’s alphabet and the newly adopted alphabets of other Turkic republics.

Distinctive Features

Turkmen exhibits most linguistic features typical of the Turkic family (*see Turkic Languages*). It is an agglutinative language with suffixing morphology, sound harmony, and a head-final constituent order. In the following, only a few distinctive features will be dealt with. In the notation of suffixes, capital letters indicate phonetic variation, e.g., *A = a/e*, *I = i/i*. Segments in round brackets only occur after consonant final stems. Hyphens are used here to indicate morpheme boundaries.

Phonology

Turkmen has, like Yakut and Khalaj, preserved Proto-Turkic long vowels in a consistent way, e.g., *a:t* ‘name’ < *a:t* (but at ‘horse’ < *at*), *dö:rt* ‘four’ vs. Turkish *dört* < *tö:rt*. The orthography does not normally mark vowel length, but *ü*: in words of Turkic origin is expressed by *üy*, e.g., *süyt* for [θü:t] ‘milk’.

Proto-Turkic *e*: is mostly represented by Turkmen *i*, which mostly corresponds to Azerbaijani *e*, e.g., *gi:č* ‘late’ (Azerbaijani *geç*, Turkish *geç*).

A striking feature of Turkmen pronunciation is the presence of the interdental fricatives θ and δ , which correspond to *s* and *z* in other Turkic languages, e.g., *θid* ‘you’ (Turkish *siz*).

As in Azerbaijani, the word-initial back velar *g*-corresponds to *q*- in other Turkic languages, e.g., *g̈i:δ* ‘girl’ (Azerbaijani *giz*, Turkish *kız*). Initial *b*- is preserved in *ber-* ‘to give’, *ba:r* ‘existing’, *ba-* ‘to go’ and *bol-* ‘to become’ (Turkish *ver-*, *var*, *var-*, *ol-*). The bilabial fricative β is used instead of labiodental *v*, e.g., *a:β* ‘hunt’ (Turkish *av*). It appears as the glide *w* between two vowels or between a liquid and a vowel. The bilabial fricative *f* is frequently replaced by the stop *p* in loans, e.g., *pikir* ‘thought’ (Turkish *fikir*).

Suffix vowels mostly assimilate to the quality of the preceding vowel. Turkmen displays both front vs. back harmony and rounded vs. unrounded harmony. The latter also includes suffixes with low vowels, e.g., *toy-do* [feast-LOC] ‘at the feast’ vs. *öy-dö* [house-LOC] ‘in the house’. Long *a:* and *e:* are not rounded; there are also other exceptions. Though the orthography represents the vowels of rather closely, rounding harmony is not consistently represented. Rounding is only expressed in high vowels and not beyond the second syllable. The tendency towards rounded low suffix vowels is also observed in languages such as Kirghiz, Altay Turkic (Altai), and Yakut.

Numerous consonant assimilations are observed, e.g., *men-ne* [I-LOC] ‘in me’, *g̈id-δan* [girl-ABL] ‘from the girl’, *yol-loš* [way-DER] ‘comrade’ (Turkish *ben-de* [I-LOC], *kiz-dan* [girl-ABL], *yol-daş* [way-DER]). They are mostly not reflected in the orthography.

In copies of loanwords, nonpermissible consonant clusters are dissolved by means of prothetic or epenthetic vowels, e.g., *uθθul* ‘chair’ < Russian *stul*, *pikir* ‘thought’ < Arabic *fikr*. In recent loanwords from Russian, these vowels are not reflected orthographically.

Grammar

The comparative degree of adjectives is formed with *-rA:K*, e.g., *kičire:k* [small-COMP] ‘smaller, rather small’ (*kiči* ‘small’). The demonstrative pronouns *bu:*, *šu:*, *ol* and *šo[l]* form a fourfold deictic system, expressing various degrees of distance (Turkish *bu*, *o* and *şu*).

The old present tense, mostly called ‘indefinite future,’ is formed with *-Ar*, e.g., *bil-er* [know-AOR] ‘will know’ (Turkish *bil-ir* [know-AOR]), *oqa:r*

[read-AOR] ‘will read’ (Turkish *oku-r* [read-AOR]). The negative marker is *-mAδ* in the third person, and *-mAr* in the other persons, e.g., *gel-mer-in* [come-NEG.AOR-1.SG] ‘I will not come’ (Turkish *gel-me-m* [come-NEG.AOR-1.SG], Azerbaijani *gel-mer-em* [come-NEG.AOR-1.SG]). A more focused present tense is formed with *-yA:r*, often contracted to *-yA*, e.g., *bil-ye:r* [know-PRES.3.SG] ‘knows’, *oqa-ya:r* [read-PRES.3.SG] ‘reads, is reading’. A few verbs exhibit contracted forms without this marker: *du:r* [stand-PRES.3.SG] ‘is standing’, *oẗi:r* [sit-PRES.3.SG] ‘is sitting’, *yaẗi:r* [lie-PRES.3.SG] ‘is lying’. These forms can be used with a converb marker to express a continuous present, e.g., *oqa:-p oẗi:r* [read-CONV AUX-PRES.3.SG] ‘is reading’.

The second-person imperatives include an unmarked singular, e.g., *gel* [come.IMP.2.SG] ‘come!’, a form expressing insistence, e.g., *gel-gin* [come-IMP.2.SG], a plural form, e.g., *gel-ij* [come.IMP.2.PL], and intensifying forms, e.g., *gel-θen-e* [come-IMP.2.SG] (singular) and *gel-θe-ηid-lä:η* [come-IMP.2.PL] (plural). The first-person plural has a special form that only refers to the speaker and the addressee, e.g., *gel-eli-η* [come-IMP.1.PL] ‘let us come’, *gel-eli* [come-IMP.1.INCL] ‘let us come (you and me)’.

The future marker *-jAK* and the intentional marker *-mAK-čI* lack personal markers, e.g., *men gel-jek* [I come-FUT] ‘I will come’ (Turkish *gel-eceğ-im* [come-FUT.1.SG]), *men yaδ-maq-č̈i* [I write-INTENT] ‘I intend to write’.

Turkmen has a postterminal (‘past’) participle marker *-An* and an intraterminal (‘present’) participle marker *-yA:n*, e.g., *bil-en* [know-POSTTERMINAL.PART] ‘having known’, *bil-ye:n* [know-INTRATERMINAL.PART] ‘knowing’. A categorical negation is formed with the participle in *-An* + possessive suffix + *yo:q* ‘non-existing’, e.g., *al-am-o:q* (<*al-an-im yo:q*) [take-POSTTERMINAL.PART-POSS.1.SG not-existing] ‘I did/do not take at all’. There is a postterminal converb marker *-(I)p*, e.g., *oyno-p* [play-POSTTERMINAL.CONV] ‘having played’. The corresponding marker of Turkish and Azerbaijani displays the uncontracted form *-(y)Ip/-(y)Ib*, e.g., Turkish *oyna-yip* [play-POSTTERMINAL.CONV].

Among the evidential markers, the inflectional suffix *-(I)p-dIr*, negated *-mAn-dIr*, forms an evidential (indirective) past, e.g., *gel-ip-dir* [come-POSTTERMINAL.CONV-EV.3.SG] ‘has evidently come’. The copula particle *eken* combines with various participles, e.g., *gel-en eken* [come-POSTTERMINAL.PART EV.PARTICLE] ‘has obviously arrived’. The copula particle *-mIš* suggests second-hand information, e.g., *gel-ip-miš-in* [come-POSTTERMINAL.

CONV-EV.3SG] ‘has reportedly come’. A presumptive intraterminal (present, imperfect) is formed with *-yA:n-dlr*, a presumptive postterminal (perfect) with *-A:n-dlr*, e.g., *bar-ya:n-nür* [go-INTRATERMINAL.PART-PRESUMP.3.SG] ‘is probably going’, *bar-an-nür* [go-POSTTERMINAL.PART-PRESUMP.3.SG] ‘has probably gone’.

A number of postverb constructions with converbs plus auxiliary verbs, *goy-* ‘to put’, *git-* ‘to go away’, *çiq-* ‘to go out’, *dur-* ‘stand’, *otur-* ‘sit’, *yör-* ‘move’, etc., express modifications of the manner in which the action denoted by the lexical verb is carried out.

Lexicon

The Turkmen vocabulary is basically of southwestern Turkic origin, though it also contains words typical of the Northwestern and Southeastern branches of Turkic. There are synonyms representing Oghuz and non-Oghuz types, e.g., *gäpî* and *işik* ‘door’, *dodaq* and *erin* ‘lip’. The vocabulary contains numerous words of Arabic and Persian origin, borrowed from Persian and representing the traditional sphere of Islamic civilization, e.g., *xat* ‘letter’, *inθa:n* ‘human being’, *ša:t* ‘glad’, *gül* ‘flower’, *irenk* ‘color’. The Turkmen conjunctions are mainly of Arabo-Persian origin, e.g., *we* ‘and’, *emma:* ‘but’. Words of Russian origin, borrowed from the 19th century on, represent phenomena of modern life, e.g., *poθyolok* ‘settlement’, *gäðyet* ‘newspaper’, *fe:рма* ‘farm’. The vocabulary contains many recent internationalisms borrowed via Russian.

Dialects

Turkmen dialects and subdialects are referred to by the names of tribes and clans. One main dialect group comprises the Teke, Yomud, Sariq, Salir, Gökleng and Ersari dialects, which are rather close to Standard Turkmen. The Teke dialect, occupying the central

area, has two subdialects, Mari and Akhal, the latter spoken in the Ashgabat region. The Yomud dialect is spoken on the southeast shore of the Caspian Sea and in the northern part of Turkmenistan. Ersari dialects are spoken in the eastern part of the country. The second main dialect group is found in the regions on and beyond the borders to Iran and Uzbekistan. These dialects are more distant from Standard Turkmen, lacking, for example, the interdental pronunciation of the sibilants *s* and *z*.

An isolated variety of Turkmen is Türkpen (Russian Trukhmen), spoken by small groups (ca. 12 000) on the lower Kuma River in the Stavropol region of Northern Caucasus. Türkpen is strongly influenced by Noghay (Nogai). Its speakers are descended from Turkmen tribes that migrated here in the 18th century from the Mangyshlak region east of the Caspian Sea. Salar, spoken in western China, seems to go back to an early Turkmen variety.

Bibliography

- Baskakov N A *et al.* (eds.) (1970). *Grammatika turkmenskogo jazyka*. Aşxabad: Akademija Nauk Turkmenskoy SSR.
- Bazin L (1959). ‘Le turkmène.’ In Deny J *et al.* (eds.) *Philologiae turcicae fundamenta 1*. Aquis Mattiacis: Steiner. 308–317.
- Clark L (1998). *Turkmen grammar*. *Turcologica 34*. Wiesbaden: Harrassowitz.
- Dulling G K (1960). *An introduction to the Turkmen language*. Oxford: Oxford University Press.
- Hanser O (1977). *Turkmen manual. Descriptive grammar of contemporary literary Turkmen. Texts. Glossary*. Wien: Verlag des Verbandes der wissenschaftlichen Gesellschaften Österreichs.
- Johanson L (2001). ‘Turkmen.’ In Garry J & Rubino C (eds.) *Facts about the world’s major languages: An encyclopedia of the world’s major languages, past and present*. New York, Dublin: The H. W. Wilson Company, New England Publishing Associates. 766–769.