THE TURKIC LANGUAGES
Arienne M. Dwyer

Speakers and Geographic Distribution

The Turkic languages are spoken across Eurasia from eastern Siberia (Yakut) to Iran (Khalaj), from China (Sarïg Yoghur) to the Ukraine (Karaim), concentrated in Central Asia, where groups of 2-20 million such as Uzbek, Kazakh, Kirghiz, and Turkmen are represented; the total number of speakers exceed 130 million. Turkish (ca. 57 million) is by far the largest group, including a population of ca. 2 million in Europe, primarily in Germany and the Netherlands. Despite their broad geographic reach, the Turkic languages have a high degree of mutual intelligibility. Languages in peripheral areas, however, tend to have many innovations due to the historical mobility of the Turkic peoples, and to contact with other language families. Turkic peoples lived in the paths of countless invasions of Eurasia, and they comprised a large part of the nominally Mongol army. This contact and mobility has rendered classification difficult.

Sound System

Most Turkic languages have eight phonemic vowels, which may be grouped according to backness, rounding, and height, such as Kirghiz:

TABLE 1. Vowel phonemes

	unrounded		rounded	
	front	back	front	back
High	i	ï	ü	u
nonhigh	e	a	ö	o

Some languages have more distinctions (e.g. Uyghur and Azerbayjani also distinguish an open e [ɛ] and a closed e [e]), and some, fewer (e.g. Uzbek and Salar, in which front/back distinctions are blurred by centralizing). Yakut, Turkmen, and Khalaj have phonemic vowel length distinctions; many other languages have long vowels in loanwords or from consonant contractions. Reduced vowels are found in Chuvash, Tatar, and Bashkir.

Turkic languages in western Eurasia contrast nonfinal consonants, especially stops, in voicing (p:b), whereas those in the east have surface contrasts based on aspiration (p:pʰ). Most languages also contrast various realizations of front:back post-palatal obstruents, though many distinctions are not phonemic; the final stop voicing contrast is often neutralized in favor of voicelessness, e.g. Turkish [at] 'name' but [adï] 'his/her/its name'. Consonants such as *f, v, ǯ,* and *ts* are atypical, though they occur in many Turkic languages due to contact. Native initial nasals and liquids are not found, except for the interrogative *ne* 'what?'.

Interactions between vowels and voiceless consonants have resulted in glottalization in South Siberia (Tuva, Tofa *aʔt* 'horse'), and preaspiration and spirantization in Inner Asia (Sarïg Yoghur *aʰt* 'horse', Uyghur *iʰkki* 'two').

Syllables tend to a CV(C) shape, maximally CV(V)(C)(C): VV is a long vowel or diphthong (Yakut *kü:s* 'force'), and the first consonant of CC must be a sonorant or fricative, as

in *türk* 'Turk'. Native syllables are either front or back, marked by both vowels and consonants (usually k/g vs. q/g/γ), e.g. *kel-* 'to come', *qal-* 'to remain'. There are many exceptions to this principle: most languages are disharmonic in loans and words with neutralized vowels.

The last stem syllable determines whether suffixes are front or back, voiced or voiceless, and in some languages, rounded or unrounded, e.g. Kirghiz *köz-gö* 'to the eye', *ečki-ler-im-ge* 'to my goats', *at-tar-ïm-γa* 'to my horses'. The Turkic languages vary widely in the extent of suffix harmony. Most assimilate at least in voicing and backness (palatal harmony), such as Tatar *čïq-tï* 'went out', *ket-te* 'left'. Some have roundness assimilation (labial harmony) as well, under more restricted conditions. In other languages such as Kirghiz and Yakut, rounding agreement also affects most low vowels, with exceptions. Some languages have a weakened harmonic system, such as Uzbek in contact with Iranian.

Word-level stress usually falls on the last syllable of native stems. Some suffixes are marked as unaccented, e.g. Turkish '*jazmidi* 'she/he did not write'. Heavy syllables attract both pitch and stress accent, such as Uyghur *dun'jäsï* 'his/her world' or '*türklär* 'Turks'.

Morphology and Syntax

Turkic has highly regular agglutinative suffixation, expressing categories for person, number, tense and so on in a strict order, e.g. Sarïg Yoghur *bar-al-γe-mes-dro* go-POT-FUT-NEG-3P.DEF 'he/she cannot go'. Many suffixes are highly productive, such as agentive *-CI*: Uyghur *iš-či* 'worker' < *iš* 'work'. Prefixing is unusual, though unproductive prefixes do exist, e.g. Persian *bi-* in Turkish *biperva* 'fearless'. The only seemingly native and productive prefixing throughout Turkic is the partial adjective reduplication: Chuvash *χup-χura* 'jet black'. Some suffixes have language-specific harmony rules; others are invariable.

Nominal (nouns, pronouns, adjectives, numerals) and verbal word classes are rigorously distinguished, and generally decline in a uniform manner. Minor morphophonemic variation does occur, such as Turkish *ben* 'I', *ban-a* 'to me'. Other word classes are indeclinable; these include adverbs, interjections, conjunctions and other particles.

Turkic nouns have neither gender nor dual. Suffixes are almost uniformly ordered thusly: Uzbek *kitob-lar-im-da* book-PL-1.POSS-LOC 'in my books'. The plural is not marked in collective or numeral expressions (Kazakh *at jaqšï* horse-good 'Horses are good', *eki at* 'two horses'). Oblique cases often take the stem *-n* after the third-person possessive, e.g. Turkmen *at-ï-n-dan* 'from his/her horse'. Within the noun phrase, there is no case/number agreement between adjectives and their head nouns, e.g. Kazakh *biyik tawlar* 'high mountains'. Many enclitic and postpositional particles have case-like functions, e.g. *üčün* 'for', *birle* 'together with'.

Nouns can generally function as substantives or adjectives (Chuvash *śută* 'brightness; bright') and many suffixes can be added to both, e.g. *baš-la-* head-LA- 'to begin', *qara-la-* black-LA- 'to blacken'. Only adjectives have comparative forms: Tatar *yaxšï-raq* 'better'.

Personal pronouns lack an inclusive/exclusive distinction; honorifics are often formed with plural pronouns, or with the reflexive pronoun in the third-person, usually *öz-i*, Turkish *kendi-si* 'she/he [himself/herself]'. Demonstrative pronouns usually have a three-way deixis, e.g. Kazakh *bul* 'this (visible)', *sol* 'that (invisible)', *osï* 'that (further away)'. Possessives are rendered with personal and demonstrative pronouns in the genitive. Inflected pronouns give rise to a number of demonstratives and interrogatives: Turkish *ne-den* what-ABL 'why?'.

Cardinal numerals are generally based on a decimal system; eleven to nineteen are additive, e.g. *on bir* ten-one 'eleven' (cf. Sarïg Yoghur and Old Turkic: *bïr yigïrmï* one-twenty 'eleven', *bïr ohdïs* one-thirty 'twenty-one'). Normally the decades from sixty to ninety are multiplicative. Ordinals take the suffix *-(I)nči*, e.g. Kirghiz *ekinči* 'second', collective numerals with *-(AG)U*, *-(I)lAn*, etc.: Kazakh *üšew* 'three together', Khakas *altolang* 'six together'.

Verbal expressions in Turkic comprise copular enclitics and regularly-suffixing verbal phrases. Nonpast copulas are generally unmarked in the third person: Tuva *ol suruqčï* 'she/he is a

student', Kazakh *bul adam jaqsï* 'this person is good', cf. *bul jaqsï adam* 'This is a good person'; when overtly marked, may take the form of nominal predicates with personal suffixes: Turkish *evdeyiz* home-LOC-1PL 'we are at home'. Negative and past copulas are always marked with particles: Turkish *ben değil* 'it isn't me', *ben değildi* (< *değil idi*) 'It wasn't me', Kazakh *men jumïssï edim* 'I was a worker'. Existence is expressed with the enclitics *bar* and *yoq*: Turkmen *kim bar kim yoq?* who-exist-who-not.exist 'who is [here] and who isn't?'.

Verbal morphology is extensive, with voice, mood, aspect, tense, and possibility/potentiality suffixes, e.g. Kazakh *kör-ïs-tïr-ïl-* 'to be shown'; with negative *-MA-*, Turkish *day-an-ïš-tïr-ïl-ma-dï* prop.up-REFL-RECIP-CAUS-PASS-NEG-SIMP.PAST-3SG 'They were not made to help each other'. Potentiality is often marked with a grammaticized al- 'take' or *bil-* 'know' as in Turkish *ver-e-bil-ir* 'she/he can give'.

Verbs fall into two classes, finite and non-finite. Finite verbs are conjugated, with markers for aspect, mood, and/or tense, and constitute independent utterances; non-finite verbs do not carry such markers, and are bound forms with conjuncting or subordinating suffixation, functioning as e.g. relative clauses or the thematic first verb in an aspectual/actional complex.

Actionality (*Aktionsart*), the manner in which an action is carried out, is typically expressed by semantically-fused verbal phrases consisting of a non-finite lexical verb and conjunctor followed by an auxiliary verb, e.g. Chuvash *ils-e pĭr-* take-go 'bring', Uyghur *išliwat-* (< *išlep yat-* work-lie) 'working'. Languages generally have one to two dozen of such grammaticalized auxiliary converbs at their disposal. In this brief summary, it is impossible to do justice to the range of morphosyntactic possibilities in combinations of tense, aspect, and modal suffixes.

One important feature of Turkic is inferentiality, where the speaker distinguishes direct from indirect experience, e.g. Uzbek *xatå qildim* 'I made a mistake' vs. *xatå qilibman* 'It seems I have made a mistake'. The indirective copulas *imiš* and *iken~eken* also exist.

Syntactic typology in Turkic is very consistent and economical. The adjunct always precedes the head; noun phrase arguments thus precede the verb, verbs come at the end, and the position just before the verb is dedicated to new information. In syntactically parallel constructions, only the last verb requires person/number/case/etc. marking: Turkish *görmüš ve duymuš-lardï.* '[They] has seen and heard [it]."

Diachronic Development

Common Turkic was theoretically the language unity the predated typological divergence into language branches; some scholars also posit an earlier Proto-Turkic language.

Old Turkic includes East Old Turkic, Old Uyghur, and Karakhanid. East Old Turkic is the oldest known Turkic runiform inscriptions on stone monuments in the Orkhon and Yenisey valleys in modern-day Mongolia from the 2nd Turkic kaghanate; the language of the early Oghuz, Kipchak, and Uyghur Turkic peoples was little differentiated at the time. Old Uyghur, flowered in the Tarim Basin Uyghur dynasty under Manichaeanism and Buddhism (9-13th c.); Karakhanid Turkic (11-12th c.), an Islamic literary language centered in Kashghar, was influenced lexically by Persian and Arabic and written in the Arabic script.

Middle Turkic reflects the increasing differentiation of the Turkic branches and the development of literary standards. In the east, Chaghatay developed as the premier pan-Central Asian Turkic literary language, written in an Arabic script. Stages include Khorezmian Turkic (13-14th c.), Early-Late Chaghatay (15-early 20th c). In the west, Oghuz Turkic included Old Anatolian Turkish (13th c. onwards), its successor Ottoman Turkish, and literary Turkmen and Azerbayjani (14th and 15th centuries onwards, respectively); Kipchak Turkic is exemplified by the 14th-century Codex Cumanicus.

Modern Turkic comprises six branches, classified on the basis of both genetic and areal-typological features. Early on, the so-called Oghur or Bulghar branch in the west split off from

Common Turkic. Oghur has *r* and *l* where Common Turkic has *ẕ* and *ʃ* in some words: Chuvash (the only surviving Oghur member) *śul* 'year', Turkish *yil*. Khalaj, spoken today in Iran, represents a further early split. Common Turkic itself has four branches: Southwestern Turkic (Turkmen, Azerbayjani, Turkish, and Gagauz); Northwestern Turkic (Kazakh, Kirghiz, Karakalpak, Noghay, Tatar, Bashkir); Northeastern (Tuvan, Khakas, Yakut), and Southeastern (Uyghur, Uzbek).

Is Turkic related to Mongolian?

A comparison of Turkic and Mongolian reveals an abundance of lexical cognates, similar suffixation and phonology. Scholars disagree whether the relationship is genetic or one of borrowing. The former theory posits a common Altaic proto-language, with three main families: Turkic, Mongolic, and Manchu-Tungusic. Altaicists pointed to regular sound correspondences, common suffixes and personal pronouns, and syntactic similarities. Proponents of the now-dominant Turko-Mongolian hypothesis asserted that such similarities were due to borrowing, citing the lack of common numerals, and similarities between Chuvash and Mongolian. Mongolian's *r* corresponding to Old Turkic *ẕ* may indicate that the Mongolian words were borrowed, likely very early on, from a *r*-type Turkic language like Chuvash: Mo. *urtu* 'long', OT *uzun*, Chuv. *vărăm*. Altaicists cite this type of example as a systematic genetic correspondence of Mo. *r* : Tkc. *ẕ* (< Proto-Turkic *ř*), and call into question the claim of unidirectional borrowing from Turkic into Mongolic. Most all scholars agree that the Mongolian and Turkic show evidence of heavy bidirectional copying at least since the intensive contact of the thirteenth-century Chinggisid Mongol empire.

 The Mongolian influence was particularly strong in Yakut, Tuva and other Siberian Turkic languages. Other early loans include Indo-European, Uralic, and Sinitic: Turkic *tümen* '10,000', Tocharian *t(u)mane*; Old Turkic *böẕ* 'linen', Greek *bussos*; O.T. *biti-* 'to write', Old Chinese *piet* 'writing brush'. In addition, Yakut shows Tungusic and Samoyedic influence, and Salar, Chinese and Tibetan elements. Uzbek, though a Southeastern Turkic language, shows the effects of contact with Tajik (Iranian) and Kipchak (Northwestern Turkic).

Arienne M. DWYER

FURTHER READING

CSATÓ, Éva Ágnes and Lars JOHANSON, eds. (1998). The Turkic Languages. London: Routledge.
ERDAL, Marcel (1991). Old Turkic Word Formation. Wiesbaden: Harrassowitz.
GOLDEN, Peter B. (1992). An Introduction to the History of the Turkic Peoples. Wiesbaden: Harrassowitz.