Appendix: Transcription, notations and abbreviations

Technical issues

This publication uses the conventions, notations, and abbreviations for morphological analysis applied in *Turkic* by Johanson (2021), *The Turkic Languages* edited by Johanson & Csató (2022), *Encyclopedia of Turkic Languages and Linguistics* (Brill) edited by Johanson, and those recommended for authors of the journal *Turkic Languages*. You can find the following guide in these publications also.

Transcriptions

Turkic languages have been written, and are written, in a variety of scripts and according to various orthographic conventions. Employing all these script systems or representing them by transliterations would be an inadequate basis for crosslinguistic comparisons in a survey such as the present one. Linguistic data are therefore normally given in a broad phonetic transcription (without brackets) that reflects typical sound segments, whether phonemic or subphonemic, e.g. Turkish *ačïḳ* 'open', Tatar *yün* 'cheap', Turkmen *soŋ* 'end', Uzbek *åt* 'horse'. Graphic forms representing official orthographies as well as cited transcriptions are given in angle brackets (chevrons), e.g. Turkish ‹açık› 'open', Tatar ‹юнь› 'cheap', Turkmen ‹soň› 'end', Uzbek ‹ot› 'horse'. Standard Turkish examples are, as a rule, rendered in their relatively well-known orthographic shapes, which are easily analyzable in phonetic terms, e.g. ‹açık› 'open'. Where precise phonetic details are needed, IPA (International Phonetic Alphabet) transcriptions are given in square brackets, e.g. Turkish [atʃɯk] 'open'. Transcriptions indicating phonologically relevant units are enclosed in slashes, e.g. Turkish/at/ 'horse'.

Broad transcription

The broad transcription largely follows traditional Turcological conventions and is similar to the epoch-making *Fundamenta* notation, applied in Deny et al. eds. (1959: XIV-XV), though it differs from it in some minor points. This transcription is chosen in order to achieve maximum uniformity without deviating too much from the actual sound structures. It does not reflect graphic differences. For instance, Bashkir orthography employs the separate graphemes ‹к› and ‹ҡ› for front *k* and back *ḳ* respectively, e.g. ‹кил-› *kil-* 'to come' and ‹ҡал-› *ḳal-* 'to stay', whereas the Tatar orthography uses ‹к› for both, namely ‹кил-› and ‹кал-›.

The symbols employed for the broad transcription are given below, with the most typical phonetic realizations indicated in IPA. Digraphs are avoided, that is, individual segments are written with separate signs. The ligatures *t͡s* and *d͡z* are, however, used to distinguish the respective affricates from the clusters *t* + *s* and *d* + *z*.

Unrounded front and near-front vowels

i = front high [i] (cardinal vowel 1)

i̯ = front near-high lax

e = front upper-mid [e] (cardinal vowel 2)

ä = front lower-mid [ɛ] (cardinal vowel 3) ~ front near-low [æ] (higher than cardinal vowel 4 [a])

Unrounded back and near-back vowels

ï = back high [ɯ] (cardinal vowel 16) ~ near-back high [ɨ] (cardinal vowel 17)

i̤ = back near-high lax

a = back low [ɑ] (cardinal vowel 5) ~ back lower-mid [ʌ] (cardinal vowel 14)

å = back low slightly labialized [ɒ] (cardinal vowel 13)

Rounded front and near-front vowels

ü = front high [y] (cardinal vowel 9)

ṳ̈ = near-front near-high lax

ö = front upper-mid [ø] (cardinal vowel 10) ~ front, lower-mid [œ] (cardinal vowel 11)

Rounded back and near-back vowels

u = back high [u] (cardinal vowel 8)

ṳ = near-back near-high lax

o = back upper-mid [o] (cardinal vowel 7) ~ back lower-mid [ɔ] (cardinal vowel 6)

Central vowels

ə = mid central lax

u̇ = rounded high central

ȯ = rounded mid central

ȧ = unrounded mid central

Labial consonants

p = strong bilabial stop [p]

b = weak bilabial stop [b]

f = strong labiodental fricative [f]

v = weak labiodental fricative [v]

ʋ = labiodental approximant [ʋ]

φ = strong bilabial fricative [ɸ]

β = weak bilabial fricative [β]

m = bilabial nasal [m]

w = bilabial glide [w]

Dental/alveolar consonants

t = strong stop [t] (on consonant strength see § 12.3.1)

d = weak stop [d]

θ = strong fricative [θ]

δ = weak fricative [ð]

s = strong fricative [s]

z = weak fricative [z]

ʦ = strong affricate [ʦ]

ʣ = weak affricate [ʣ]

n = nasal [n]

l = voiced lateral approximant [l]

r = rhotic: vibrant, trill [r] ~ tap/flap [ɾ]

Postalveolar-palatal consonants

c = strong palatal plosive

č = strong affricate [ʧ]

ɟ = weak palatal plosive

ǰ = weak affricate [ʤ]

š = strong fricative [ʃ]

ž = weak fricative [ʒ]

ś = strong alveolo-palatal fricative [ɕ]

ź = weak alveolo-palatal fricative [ʑ]

ʂ = voiceless retroflex fricative [ʂ]

ʐ = voiced retroflex fricative [ʐ]

y = palatal glide [j]

ń = palatalized nasal [ɲ]

Velar-postvelar consonants

k = front strong stop, typically [k]

ḳ = back strong stop, typically [q]

g = front weak stop, typically [g]

ġ = back weak stop, typically [G]

χ = strong fricative [χ] or [X]

ɣ = weak fricative [ɣ] or [ʁ]

ŋ = nasal [ŋ] or [ɴ]

Glottal consonants

h = glottal fricative [h]

ʾ = glottal stop [ʔ]

Diacritics

The broad transcription indicates relatively front and back articulation of certain consonants. Though the differences may be more or less audible in the various Turkic languages, they play crucial and systematic roles in the systems of sound harmony. Certain consonants with a relatively back articulation may be indicated with a superscript or subscript dot, i.e. *ḳ, ġ, ḷ, ɣ̇, χ̇, ŋ̇*. Notations of this kind are required in discussions of sound harmony systems. Corresponding consonants with a relatively front articulation can be written with the dotless signs *k, g, l, ɣ, χ, ŋ*. If necessary, a superscript or subscript right arrowhead may be used for front consonants, e.g. *k̛, ġ̛, l̛*. More precise [±front] differences can be indicated with narrower phonetic signs. For instance, in some West Oghuz varieties, the front *k̛* has a very advanced articulation that can be represented by means of the IPA symbol [c]. In order not to overburden the transcriptions, normally only the [±front] distinctions *k* vs. *ḳ* and *g* vs. *ġ* are indicated in this volume.

In traditional Turcological transcriptions, the strong velars *k* and *ḳ* are written ‹k› and ‹q›, respectively. The weak velars *g* and *ġ* are mostly written as ‹g›, and the fricatives *ɣ* and *ɣ̇* as ‹ɣ›. Recent transcription practices for East Old Turkic, allegedly based on phonemic analyses, employ ‹k› for *k*, *ḳ* and ‹g› for *g*, *ġ, ɣ, ɣ̇*.

In transliterations of the East Old Turkic runiform script (§ 19.2), the numerical symbol ^{1} may be used for consonants associated with a latent back vowel, ^{2} for consonants associated with a latent front vowel, and ^{0} for consonants associated with back or front latent vowels.

Long consonants are marked with a triangular colon, e.g. *t:*. When occurring at morpheme boundaries they are written as a sequence of two consonants, e.g. *tt*.

Small caps (short uppercase letters) indicate weakly voiced weak obstruents, e.g. *B*, *D*.

Palatalization of consonants is indicated by a superscript acute sign, e.g. *d́, ǵ, ǰ, ḱ, ĺ, ń, ś, t́, ź, ś*.

A right-sided apostrophe (’) designates ejective articulation, e.g. *t’, k’, ḳ’*.

A raised letter may indicate a prothetic glide preceding word-initial vowels, e.g. Kazakh ^{y}*ekị* ‘two’, and evanescent segments, e.g. *b^{i}r* ‘one’. Phonologically relevant strong aspiration is marked with a raised left- or right-sited ^{h}, e.g. Dukhan *a^{h}t* ‘horse’, *p^{h}ar* ‘tiger’. Phonologically irrelevant weak aspiration is left unmarked. A weak glottal onset of word-initial vowels is mostly left unmarked or, if necessary, marked by a left-sited raised glottal stop sign ^{ʔ}. This sign also indicates a glottal element occurring between a vowel and a following obstruent, e.g. Tuvan *a^{ʔ}t* ‘horse’ (preglottalization).

The transcriptions in this volume do not make use of the symbol “X”, which is often employed in Turcological literature to indicate fourfold harmonic alternations allegedly comprising the high vowels *i*, *ï*, *ü*, *u*, in reality often the lax vowels *i̱*, *ï̱*, *ü̱*, *u̱*.

Vowel length is indicated by a right-sited triangular colon instead of a macron, e.g. *a:* instead of ‹ā›. Possible half-long vowels may be indicated by a right-sited raised dot, e.g. *a·*. If necessary, a superscript breve (˘) can be used to designate extra-short vowels. Diphthongs are written as sequences of two vowel signs.

A superscript tilde (~) designates nasalization, e.g. *ã* (nasal vowel), *ỹ* (nasalized palatal glide).

Other conventions

An acute accent (´) over a vowel marks a syllable carrying high pitch, e.g. Turkish ‹köy|dé› ‘in the village’.

A grave accent sign (`) can denote a light pitch accent, e.g. ‹lo'kanta|làr› ‘restaurants’.

A high vertical stroke (') in front of a syllable indicates dynamic stress, e.g. Turkish ‹'belki› 'maybe'.

A low vertical stroke (ˌ) in front of a syllable may indicate slight dynamic stress, e.g. Turkish ‹ˌanné› 'mother'.

A long vertical stroke (|) may divide segments in orthographic word forms, e.g. ‹köy|de› 'in the village'.

Curly brackets { } are used for morphophonemic formulas, e.g. {-DIr} (causative marker).

Mathematical angle brackets ⟩ are used for glosses, e.g. *at-lar* ⟨horse-PL⟩ 'horses'.

The sign < means 'developed from', whereas > means 'developed into'.

Simple arrows are used for morphological derivation. Thus, ← means 'derived from', whereas → means 'derived as'.

Double arrows are used for copied (borrowed) elements. Thus, ⇐ means 'copied from', whereas ⇒ means 'copied as'.

The tilde ~ between two elements denotes variation.

In descriptive contexts, an asterisk (*) indicates that the following construction is regarded as unacceptable, mostly ungrammatical. In discussions on diachronic issues, asterisks are used for reconstructed rather than actually attested items. Two asterisks are used to mark hypothetical Proto-Turkic segments, e.g. *k*.

A lowered index sign ᵢ denotes coreferentiality.

Hyphens are placed to the right of verb stems, e.g. *al-* 'to take'.

Contrary to the orthographic practices in the respective languages, hyphens are often used to show the segmentation of complex forms, indicating the boundaries between constituent segments, usually, but not always, morphemes, e.g. Turkish *at-tan* ‹at|tan› 'from the horse', Chuvash *kil-ä-ɣän* ‹килекен› 'coming'.

Example sentences in texts are written with an initial capital letter but without a final full stop, e.g. Turkish ‹Geldim› 'I came'.

The paragraph sign § is used to refer to places in the main text.

In translations, X is used as a shorthand for the 3SG personal pronouns 'he', 'she', 'it', 'him', 'her'. Other abbreviations are: (S) subject, (DO) direct object, (IO) indirect object, (P) predicate. On 'S' ('intransitive subject'), 'A' ('transitive subject'), and 'O' ('transitive object') see § 51.2.

Names of languages and their varieties are normally not abbreviated. Some of the exceptions are EOT = East Old Turkic and WOT = West Old Turkic.

Morphophonemic notations

Morphophonemic formulas, given in curly brackets, summarize the possible realizations of bound morphemes. Upper-case letters indicate elements alternating according to phonological rules of sound harmony and consonantal voice variation. They are used in archimorphemic forms of suffixes (Johanson & Csató 1998: xx-xxi). If necessary, the number of alternants may be indicated by raised numerical symbols.

Table 1. Examples of morphophonological notations

{C}	consonant
{V}	vowel

{°}		vowel of unknown quality
{Ø}		zero
{W}		alternation of high and low rounded vowels
{A}		alternation of low vowels
	{A²}	twofold alternation of low unrounded vowels, *a*, *ä*
	{A⁴}	fourfold alternation of low unrounded and rounded vowels, *a*, *ä*, *o*, *ö*
{I}		alternation of high vowels
	{I²}	twofold alternation of high unrounded vowels, *i*, *ï*
	{I⁴}	fourfold alternation of high unrounded and rounded vowels, *i*, *ï*, *ü*, *u*
{U}		alternation of high rounded vowels, *ü*, *u*
{Ị}		alternation of near-high lax vowels
	{Ị²}	twofold alternation of near-high unrounded lax vowels, *ị*, *ï*
	{Ị⁴}	fourfold alternation of near-high unrounded and rounded lax vowels, *ï*, *ị*, *ụ*, *ü*
{O}		alternation of rounded low vowels, *o, ö*
{B}		alternation of labials
{D}		alternation of dentals
{L}		alternation of *l*, *r*, *n*
{J̌}		alternation of postalveolar-palatals
{G}		alternation of weak front velars
{Ġ}		alternation of weak back velars
{K}		alternation of strong velars

Further conventions in morphophonemic notations:

Deverbal suffixes are marked with a minus sign, e.g. the Turkish 3SG voluntative suffix {-sỊn}.

Other suffixes are marked with a plus sign, e.g. the Turkish locative marker {+DA}.

Brackets of the type ( ) are used for segments that occur in certain environments and are lacking in others, depending on the stems to which they are attached.

Thus, a bracketed initial vowel sign indicates a vowel that occurs after consonant-final stems and is absent after vowel-final stems, e.g. in the Gagauz 1SG possessive marker {+(Ị)m}, realized as *iš-ịm* 'my work' and *boba-m* 'my father'.

A bracketed initial consonant sign indicates a segment that occurs after stem-final vowels and is absent after stem-final consonants, e.g. in markers such as the Turkish dative suffix {+(y)A} = ‹-a›, ‹-e›, -‹ya›, ‹-ye›, or the genitive marker {+(n)Ịn}, realized as ‹ev|in› ⟨house-GEN⟩ and ‹gece|nin› ⟨night-GEN⟩.

A bracketed final consonant sign indicates a consonant that occurs before case suffixes and is absent in other environments, e.g. 3SG possessive markers of the type {+(s)Ị(n)}, as in the Turkish dative forms ‹baba|sın|a› 'to his/her father' and ‹ev|in|e› 'to his/her house'.

A bracketed initial zero sign (Ø) indicates that the stem-final vowel drops when the marker is added, as in the Chuvash intraterminal marker {-(Ø)At}, e.g. *Vul-at* 'X reads' ← *vula-* 'to read'.

A bracketed colon (ː) indicates that the length of the suffix-initial vowel is preserved, e.g. the Turkmen participle marker {-An}//{-(ː)n}, with forms such as *oḳoːn* 'having read' ← *oḳa-* 'to read'. See § 20.8 on the term 'participle'.

If it is impossible to indicate postconsonantal and postvocalic alternants in one formula of this kind, double slashes can be used to separate the alternants.

Thus, {-A//-y} indicates alternation of a non-high unrounded vowel and a palatal glide. A bracketed upright arrow (↑) indicates raising of a non-high stem-final vowel, as in the Tatar intraterminal marker {-A//-(↑)y}, e.g. *ḳarï-y* 'looking' ← *ḳara-* 'to look'.

Formulas such as the one employed for the Turkmen participant nominal marker {-An//-ːn} indicate that, in postvocalic alternants, the stem-final vowel and the suffix-initial vowel merge into a long vowel, e.g. *oḳoː-n* 'having read' ← *oḳo-* ‹oka-› 'to read'.

Formulas such as the one employed for the Turkmen aorist marker {-Ar//-(Ø)Aːr} or the Altay 1PL exclusive voluntative marker {-AlḶK//-(Ø)AːlḶK} indicate that, in postvocalic alternants, the stem-final vowel is dropped and the suffix-initial vowel is lengthened.

The formula {-AːlḶ//-(Ø)AːlḶ} as the one employed for the corresponding Tofan suffix indicates that, in postvocalic alternants, the stem-final vowel is dropped whereas the length of the suffix-initial vowel is preserved.

Abbreviations in glosses as recommended in the journal Turkic Languages and additions from the authors in this volume

A	converb in A
ABL	ablative
ACC	accusative
ACC.DAT	accusative-dative
ACT	actional marker
ADJ	adjective marker
ADV	adverb
AFF	affirmative
AGR	subject person-number agreement marker
AGR.POSS	agreement marker of the possessive (non-enclitic) type
AGR.PRON	agreement marker of the pronominal (enclitic) type
AN	action nominal
AOR	aorist
AUX	auxiliary verb
B	converb in B
CAUS	causative
COLL	collective
COM	comitative, comitative-sociative

COMP	comparative
COND	conditional
CONJ	conjunction
CONV	converb
COOP	cooperative-reciprocal
COORD	cooordinative
COP	copula
COP.PTCL	copular particle, e.g. {+DIr}
DAT	dative
DAT.LOC	dative-locative
DER	derivational marker
DIM	diminutive
DIR	directive
DIST	distance
DIST.COP	distant copula
DISTR	distributive
DU	dual
EMOT	emotive
EMPH	emphasis
EQU	equative
EVID	evidential
EVID.COP	evidential copula
EX	existential
EXCL	exclusive
FEM	feminine
FILL	filler
FOC	focal
HFOC	high focal
GEN	genitive
HAB	habitual
HYP	hypothetical
HYP.COP	hypothetical copula
IMP	imperative
INCL	inclusive
INDEF	indefinite (pronoun)

INF	infinitive
INSTR	instrumental
INT	intentional
INTRA	intraterminal
INTRA.COP	intraterminal copula
INTRJ	interjection
INTS	intensifier
ITR	intransitive
IZ	izafet marker
JUNCT	junctor
LFOC	low focal
LOC	locative
LOC.ABL	locative-ablative
MOM	momentaneous
N	noun
NEC	necessitative
NEG	negation
NFOC	non-focal
NOM	nominative
OBJ	object
OBL	oblique
OPT	optative
OPT.COP	optative copula
ORD	ordinal numeral
P	person
1	first person
2	second person
3	third person
PASS	passive
PT	partitive
PL	plural
PL.COP	plural copula
PN	participant nominal
POL	polite
POSS	possessive

POST	postterminal
POSTP	postposition
POT	potential
PREP	preposition
PRIV	privative
PROL	prolative
PRO	prospective (“future”)
PRON	pronoun
PROPR	propriative
PTCL	particle
PURP	purposive
R	Russian material inserted in Dolgan speech
REC	reciprocal
RED	reduplication
REFL	reflexive
REL	relational suffix {+KI̩(n)}
Q	question particle
SG	singular
SIM	similative
SUP	superlative
+T	transformative, transformativizing suffix
-T	nontransformative
TERM	terminal
+TF	finitransformative
+TI	initiotransformative
TOP	topic marker
TR	transitive
TRANS	subject transcendence
VBZ	verbaliser
VN	verbal nominal
VOC	vocative
VOL	voluntative