

## EESTI JA LÄHISUGUKEELTE MURDETRANSKRIPTSIOON UNICODE'i STANDARDIS

Eesti ja lähisugukeelte murdetranskriptsioon rajaneb soome-ugri transkriptsioonil (SUT; vt ülevaateid: Ariste 1980, Pajusalu jt 2002, Vende 1967). Kasutatavad märgid võib väliskuju põhjal jagada kolme rühma: 1) (puhtad) alustähed, 2) iseseisvad sümbolid, 3) kombineeruvad diakriitikud. Alustähtede ja kombineeruvate diakriitikute ühendid annavad neljanda rühma: diakriitikutega tähed (kombinatsioonid). Seda rühma saab omakorda jagada kaheks: a) aluskombinatsioonid (tähed, mis murdetranskriptsioonis võib võrdsustada alustähtedega, nt *ä, ö, ü, š*), b) muud keerulised kombinatsioonid (lisanduvad kvantiteedi- jm diakriitikud).

### 1. Alustähed

Sellesse rühma kuuluvad lisamärkideta ladina alustähestiku väiketähed *a, b, (c), d, e, f, g, h, i, j, k, l, m, n, o, p, r, s, t, u, v, w, z* ning lisaks alljärgnevad tähed, mis jagunevad põhiliselt nelja rühma: 1) kreeka ja vene tähed, 2) kapiteelid, 3) kummuli tähed, 4) ülatähed.

<i>Täht</i>	<i>Unicode</i> <sup>1</sup>	<i>Nimetus; märkusi (M)</i> <sup>2</sup>
β	03B2	kreeka beeta; ahtus- <i>b</i>
γ	03B3	kreeka gamma; ahtus- <i>g</i>
δ	03B4	kreeka delta; ahtus- <i>d</i>
ε	03B5	kreeka epsilon; kõrge <i>ä</i> . NB! ei sobi glüüf ε. M1
η	03B7	kreeka eeta; <i>ng</i> -häälik. Vt M2
κ	03BA (03F0?)	kreeka kapa; imi- <i>k</i> . NB! ei sobi glüüf κ. M1
π	03C0	kreeka pii; imi- <i>p</i>
ρ	03C1 (03F1?)	kreeka roo; uvulaarne <i>r</i> . NB! ei sobi glüüf ρ. M1
τ	03C4	kreeka tau; imi- <i>t</i>
φ	03C6	kreeka fii; bilabiaalne <i>f</i> . NB! ei sobi glüüf φ. M1
χ	03C7	kreeka hii; velaar-palataalne ahtushäälik. IPA [x]
ψ	03C8	kreeka psii; <i>ptr</i> -häälik
Л	043B	kirillitsa <b>l</b> ; vene <i>l</i> -häälik
A	1D00	kapiteel- <b>a</b> ; helitu <i>a</i> . Vt M3
B	0299	kapiteel- <b>b</b> ; poolheliline <i>b</i>
D	1D05	kapiteel- <b>d</b> ; poolheliline <i>d</i>

<sup>1</sup> Unicode'i kood 16ndkoodina. Võimalikud alternatiivid on sulgudes, vt ka märkusi.

<sup>2</sup> Tähe nimetus vastab põhilises Unicode'i standardis antule, see lähtub tähe väliskujust, mitte foneetilisest tähendusest. Järgnevad vastava hääliku lühikirjeldus (ei ole ammendav) ja viited märkustele tabeli järel.

E	1D07	kapiteel- <b>e</b> ; helitu <i>e</i>
G	0262	kapiteel- <b>g</b> ; poolheliline <i>g</i>
I	026A	kapiteel- <b>i</b> ; helitu <i>i</i>
L	029F	kapiteel- <b>l</b> ; helitu <i>l</i>
M	1D0D	kapiteel- <b>m</b> ; helitu <i>m</i>
N	0274	kapiteel- <b>n</b> ; helitu <i>n</i>
R	0280	kapiteel- <b>r</b> ; helitu <i>r</i>
U	1D1C	kapiteel- <b>u</b> ; helitu <i>u</i>
V	1D20	kapiteel- <b>v</b> ; helitu <i>v</i>
Z	1D22	kapiteel- <b>z</b> ; poolheliline <i>z</i>
æ	00E6	ladina <b>æ</b> ; ä ja ε vaheline häälik. Kursiivis <i>æ</i>
Ʒ	0292	ež-täht; dz-afrikaat. SUTs fakultatiivne
o	0254	lahtine <i>o</i> (= IPA)
ɶ	0250	kummuli <b>a</b> ; redutseeritud <i>a</i> . Kursiivis <i>ɶ</i> . M4
ɷ	0259 (01DD?)	kummuli <b>e</b> ; redutseeritud <i>e</i> . M5
ɸ	1D09	kummuli <b>i</b> ; redutseeritud <i>i</i> . M6
<sup>a</sup>	1D43	üla- <b>a</b> ; ülilühike <i>a</i> . M7

### Märkused.

1. Mõned kreeka tähed (epsilon, kapa, roo, fii) varieeruvad oma väliskujult. Unicode'i tabelis on alternatiivsetena antud kreeka kapa sümbol (03F0) ja kreeka roo sümbol (03F1), mida nimetatakse tehnilisteks sümboliteks. Nende põhikuju (glüüf) langeb kokku SUT märgiga. Samas epsilon ja fii puhul langevad SUT märgiga kokku põhitähed; tehniliste sümbolite (fii sümbol 03D5 ja epsilon sümbol 03F5) põhikujud on teistsugused. Loogiline oleks siiski võrdsustada kõik kõnealused SUT märgid kreeka põhitähestiku tähtedega ja kujundada fondis need tähed nii nagu SUTs tavaks.
2. Ariste (1980: 6) nimetab seda selgelt kreeka täheks. Alternatiiv oleks selle hääliku tähistus võrdsustada vastava IPA sümboliga  $\eta$  (014B).
3. Kapiteelide loetelu pole täielik, kuid peaks praktikas piisama. Ariste (1980: 7) seostab vastavate väiketähtedega veel järgmised kapiteelid:  $\beta > \text{Б}$  (...),  $\gamma > \Gamma$  (1D26),  $\delta > \text{Д}$  (03D1),  $\epsilon > \Sigma$ , (...)  $\eta > \text{И}$  (...),  $\rho > \text{Р}$  (1D29),  $\lambda > \text{Л}$  (1D2B),  $\alpha > \text{Æ}$  (1D01),  $\sigma > \text{С}$  (1D10). Nagu näha, ei ole kõigil kapiteelidel Unicode'i koode (või ei ole neid õnnestunud leida).
4. Kummuli **a** kood on valitud tähenimetuse põhjal. IPA-s on *a* ja *ɑ* eraldi tähed (viimast nimetatakse ladina alfaks, kood 0251), vastavalt on eri koodiga kummuli *a* ( $\mathfrak{a}$ , 0250) ja kummuli ladina alfa ( $\mathfrak{A}$ , 0252). SUT seisukohalt tuleks „ladina alfa“ paar kõrvale jätta.
5. Kummuli **e** kood on võetud IPA plokist, alternatiiv oleks kasutada ladina tähestikuploki koodi 01DD (= aserbaidžaani ja tatari *ä*).
6. Kummuli tähtede loetelu pole täielik, kuid peaks praktikas piisama.
7. Ülatähtede täielikku loetelu ei ole võimalik anda. Neid saab küll tekitada standardsete tekstitöötlusvahenditega, ent Unicode pakub välja võimaluse ka neid kodeerida. Omaette koodiga on nt ülatähed <sup>a</sup> (1D43), <sup>e</sup> (1D49), <sup>h</sup> (02B0), <sup>i</sup> (2071), <sup>j</sup> (02B2), <sup>k</sup> (1D4F), <sup>m</sup> (1D50),

° (1D52), ° (1D53), ° (1D56), ° (02B3), ° (02E2), ° (1D57), ° (1D58), ° (1D5B)

## 2. Iseseisvad sümbolid

Unicode eristab omaette tähekohaga teisendtähti (*spacing modifier letters*) ja sümboleid päristähtedest ja kombineeruvatest diakriitilistest märkidest. Viimaseid iseseisvalt ei kasutata, vaid üksnes alustähega koos.

?	0294	kõrisulghäälikutäht. = IPA
˘	02BE	parem ülalpoolring; katkehäälik (liivis)
˙	02BF	vasak ülalpoolring; hõngus (aspiratsioon)
◌◌	02F3	aladõõri teisendtäht; vokaali järel märgib labialisatsiooni
◌◌	02EF	alanool alla; märgib redutseeritud vokaali nt <i>o</i> puhul
◌◌	02DB (?)	pöördsedii teisendtäht; silbipiir
◌◌	203F	ühenduskaar; sõnade foneetiline liitumine
◌◌	02FD	riiulmärk; hääliku ülekandumine teise sõnasse
◌◌	00B7	keskpunkt; pearõhk
◌◌	003A	kaksikpunkt (koolon); kaasarõhk

## 3. Kombineeruvad diakriitikud

Neid märke saab kasutada üksnes koos alustähega. Unicode'i standardi kohaselt kirjutatakse enne alustäht, seejärel diakriitik. Kui diakriitikuid on mitu, siis on nende standardne järjestus tähe poolt vaadates seestpoolt väljapoole ja ülalt alla. St kõige enne kirjutatakse märk, mis tuleb vahetult tähe kohale, seejärel (vajaduse korral) märk, mis tuleb eelmise peale, seejärel tähe keskele lisatavad märgid, lõpuks tähe alla lisatavad märgid, alustades sellest, mis on alustähele ligemal. Diakriitilised märgid allpool on Unicode'i koodide järjestuses.

◌◌	0300	graavis; poolpikk häälik
◌◌	0301	akuut; palatalisatsioon
◌◌	0302	tsirkumfleks; ülipikk häälik
◌◌	0303	tilde; tugev nasalisatsioon
◌◌	0304	makron; pikk häälik
◌◌	0306	kaar; harilikust lühem häälik
◌◌	0307	ülalpunkt; vahepealne vokaal ( <i>a~ä, e~i, o~u, u~ü</i> )
◌◌	0308	täpid (treema); eespoolne vokaal
◌◌	030A	üladõõr; labiaalsus (nt <i>å</i> )
◌◌	030C	haak: <i>č, š</i> ja <i>ž</i>
◌◌	0311	look; (vaja siis, kui alakaarega täht keerata kummuli)
◌◌	031C	vasak alalpoolring; nõrk palatalisatsioon (koronalisatsioon)

.	0323	alapunkt; (vaja siis, kui ülapunktiga täht keerata kummuli)
..	0325	alatäpid; (vaja siis, kui täppidega täht keerata kummuli)
◦	0325	alasõõr; iseseisvat silpi märkiv häälik
,	0329	alapüstjoon; interdentaalne hääldus
˘	032C	alahaak; harilikust madalam moodustus
ˆ	032D	alatsirkumfleks; harilikust kõrgem moodustus
˘	032E	alakaar; keskvokaal
ˆ	032F	alalook; (poolvokaali vananenud tähis)
˜	0330	alatile; nõrk nasalisatsioon
˘	0337	läbiv lühike kaldkriips: <i>l</i>
˘	0354	alanool vasakule; harilikust tagapoolsem moodustus
˘	0355	alanool paremale; harilikust eespoolsem moodustus
˘	20D0	kolmveerandpikkusmärk; Unicode'is on lähim vaste nn vasakharpuun

#### 4. Aluskombinatsioonid

Murdetranskriptsiooni seisukohalt võib järgmisi alustähe ja diakriitiku kombinatsioone käsitada aluskombinatsioonidena, sest neile lisanduvad sageli muud diakriitikud (välde, palatalisatsioon jm). Unicode'is on nad enamasti olemas valmistähtedena (*precomposed characters*).

ä	00E4	<b>a</b> täppidega
å	00E5	<b>a</b> ülasõõriga; <i>o</i> -poolne <i>a</i>
č	010D	<b>c</b> haagiga; SUTs <i>tš</i> -afrikaadi fakultatiivne tähistus
ç	00E7	<b>c</b> sediiga; nn <i>ich</i> -häälik
ĕ	1E19	<b>e</b> alatsirkumfleksiga; kõrgem <i>e</i>
ē	0065+032E	<b>e</b> alakaarega; eesti <i>õ</i>
ĕ	0065+0354	<b>e</b> alanoolega vasakule; <i>õ</i> -lähedane <i>e</i>
ĭ	0069+032E	<b>i</b> alakaarega; venepärane <i>õ</i> ( <i>bi</i> )
ĭ	0069+032C	<b>i</b> alahaagiga; madalam <i>i</i>
ł	0142	<b>l</b> läbiva lühikese kaldkriipsuga; <i>o</i> -varjundiga <i>l</i>
ö	00F6	<b>o</b> täppidega
ȯ	006F+032D	<b>o</b> alatsirkumfleksiga; kõrgem <i>o</i>
ō	00F6+032D	<b>o</b> täppide ja alatsirkumfleksiga; kõrgem <i>ö</i>
ö	0254+0308	lahtine <b>o</b> täppidega; lahtine <i>ö</i>
š	0161	<b>s</b> haagiga

ų	0075+032C	u alahaagiga; madalam u
ü	00FC	u täppidega
ÿ	00FC+032C	u täppide ja alahaagiga; madalam ü
ž	017E	z haagiga

## 5. Muud keerulised kombinatsioonid

Järgnev loetelu sisaldab muid diakriitikutega kombinatsioone. Esimeses kahes tulbas on diakriitikud ja nende kombinatsioonid (Unicode'i standardses kodeerimisjärjestuses, vt 3. osa); kolmandas on loetletud alustähed ja -kombinatsioonid, millele neid võidakse lisada. Kui kogu kombinatsioon on Unicode'is olemas valmistähena, antakse sulgudes selle kood.

`	0300	a (00E0), ç, e (00E8), f, h, i (00EC), k, l, m, n, o (00F2), p, r, s, t, u (00F9), v, ε, η, λ, ä, ě, ě, ĭ, š, ö, ŏ, ü (01DC). M8
´	0301	c, d, f, g, h, k (1E31), l (013A), m (1E3F), n (0144), p (1E55), r (0155), s (015B), t, v, z, η, B, D, G, Z, č, š, ž
´+	0301+0300	h, k, l, m, n, p, r, s, t, v, D, š
´+	0301+0304	h, k, l, m, n, p, r, s, t, v, D, š
´+	0301+0306	h, k, l, m, n, p, r, s, t, v, D, š
^	0302	a (00E2), e (00EA), i (00EE), o (00F4), u (00FB), ε, æ, ɔ, ä, ě, ě, ĭ, ĭ, ö, ŏ, ɔ, ɔ, ɔ, ü, ü. M8
~	0303	...
ˉ	0304	a (0101), e (0113), f, h, i (012B), k, l, m, n, o (014D), p, r, s, t, u (016B), v, η, λ, ε, æ (01E3), ɔ, ä (01DF), ě, ě, ĭ, ĭ, š, ö (022B), ŏ, ŏ, ɔ, ü (01D6), ɔ, ü. M8
˘	0306	a (0103), ç, e (0115), f, h, i (012D), k, l, m, n, o (014F), p, r, s, t, u (016D), v, η, λ, ä, ě, ĭ, ö, ŏ, š, ü. M8
·	0307	a (0227), e (0117), o (022F), u
˘	030C	z, ɜ (01EF)
ˆ	0311	ə, ɪ
ˆ	031C	k, l, n, r, s, t
.	0323	ɐ, ə, o, n, ö
..	0324	ɐ, o, n
◦	0325	l, m, n, r
ˆ	0329	D, s, t
˘	032C	o, e, ö
ˆ	032D	ä, õ
ˆ	032F	a, e, i, u, ü

~	0330	...
<	0354	i, l, ä, ü
>	0355	a, l, o
^	20D0	a, e, i, o, u, ε, æ, ɔ, ä, ɛ, ɶ, ɷ, ö, ɔ̥, ɔ̥̥, ɔ̥̥̥, ɷ̥, ü, ȳ. M8

### Märkused.

8. Kui diakriitik liitub *i*-tähe peale, siis jääb SUTs alles ka *i* punkt, mida tuleb fondi kujundamisel arvestada ( $\bar{i}$ ,  $\grave{i}$  jne). Ladinatähealiste keelte tavaortograafias jäetakse *i* punkt neil juhtudel ära ( $\bar{i}$ ,  $i$ ).

### Allikad

**Ariste, P. 1980.** *Foneetilise transkriptsiooni ja gooti kirja harjutusi*. Tartu.

**Pajusalu, K.; Hennoste, T.; Niit, E.; Päll, P.; Viikberg, J. 2002.** *Eesti murded ja kohanimed*. Tallinn, lk 298–301.

**Vende, K. 1967** = К. Венде. *Финно-угорская транскрипция (ФУТ) в историческом аспекте и в сопоставлении с международным фонетическим алфавитом (МФА)*. Таллин.