

Sõnarõhk keeleinfos ja kõnesünteesis

Liisi Piits, Heete Sahkai, Meelis Mihkla, Indrek Hein, Liis Ermus, Hille Pajupuu

Rakenduslingvistika kevadkonverents 2025

Hääldusinfo ja kõnesüntees EKI ühendsõnastikus

☒ märkidega NB! Siin on esitatud häälduskuju, mitte kirjapilt.

MUUTTÜÜP 22e

<https://sonaveeb.ee/search/unif/est/dsall/tall/2/est>

kääne	ainsus	mitmus
nimetav kes? mis?	t`al'l 	tal'li[d 
omastav kelle? mille?	tal'li 	t`al'li[de  Tehishääli
osastav keda? mida?	t`al'li 	t`al'le  ~ t`al'li[sid 
lühike sisseütlev kellesse? millesse?	t`al'li 	
sisseütlev kellesse? millesse?	tal'li[sse 	t`al'li[desse  ~ tal'le[sse 

Hääldusmärke arvestav üksiksõnade süntees

Treeningkorpus:

- Naishääl 31215 sõnavormi, kokku 10h 36min
- Sõnades on märgendatud **palatalisatsioon, III välde, ebareeglipärane rõhk, liitsõnapiir**

Merlini kõnesünteesitööriistad

Üksiksõnade süntees kuulatav siin: elo.eki.ee/yksiksona



Eelised:

- Jäljendab üksiksõnade prosoodiat
- Suudab eristada minimaalpaare, mida ortograafia ei erista
- Võimaldab sõnade hääldust korrigeerida (sõna **vann** näide)

Üksiksõnade süntees: palatalisatsioon

Palataliseerimata	tuvastusprotsent	Palataliseeritud	tuvastusprotsent
<i>kott</i>	100%	<i>kot't</i>	88%
<i>jutt</i>	100%	<i>jut't</i>	75%
<i>nut</i>	100%	<i>nut't</i>	75%
<i>patt</i>	100%	<i>pat't</i>	100%
<i>tall</i>	100%	<i>tal'l</i>	100%
<i>Mall</i>	100%	<i>mal'l</i>	100%
<i>puss</i>	100%	<i>pus's</i>	100%
<i>kann</i>	100%	<i>kan'n</i>	100%
<i>konn</i>	100%	<i>kon'n</i>	100%
<i>palk</i>	100%	<i>pal'k</i>	100%
<i>tulp</i>	100%	<i>tul'p</i>	88%
<i>müts</i>	75%	<i>müt's</i>	100%
<i>pat</i>	88%	<i>pat's</i>	100%
<i>mulk</i>	100%	<i>mul'k</i>	100%
<i>salv</i>	100%	<i>sal'v</i>	100%
<i>sulg</i>	100%	<i>sul'g</i>	100%

Estonian isolated-word text-to-speech synthesiser

Kiissel, Indrek; Piits, Liisi; Sahkai, Heete; Hein, Indrek; Ermus, Liis; Mihkla

Proceedings of the NoDaLiDa/Baltic-HLT, 2025



Üksiksõnade süntees: välde

II vältes	tuvastusprotsent	III vältes	tuvastusprotsent
<i>paksu</i>	100%	<i>p` aksu</i>	100%
<i>vitsa</i>	100%	<i>v` itsa</i>	100%
<i>kommi</i>	100%	<i>k` ommi</i>	100%
<i>sanga</i>	100%	<i>s` anga</i>	100%
<i>saia</i>	100%	<i>s` aia</i>	100%
<i>heina</i>	100%	<i>h` eina</i>	100%
<i>koera</i>	100%	<i>k` oera</i>	100%
<i>kuuli</i>	100%	<i>k` uuli</i>	100%
<i>prints</i>	100%	<i>pr` intsi</i>	100%
<i>katke</i>	100%	<i>k` atke</i>	100%
<i>maitse</i>	88%	<i>m` aitse</i>	100%
<i>mulje</i>	100%	<i>m` ulje</i>	100%
<i>arve</i>	100%	<i>` arve</i>	100%
<i>võime</i>	100%	<i>v` õime</i>	100%
<i>hoone</i>	100%	<i>h` oone</i>	100%
<i>rooste</i>	100%	<i>r` ooste</i>	100%

Estonian isolated-word text-to-speech synthesiser

Kiissel, Indrek; Piits, Liisi; Sahkai, Heete; Hein, Indrek; Ermus, Liis; Mihkla

Proceedings of the NoDaLiDa/Baltic-HLT, 2025



Üksiksõnade süntees: rõhk

Rõhu määramisel lähtutakse ühendsõnastiku märgendusest ja lisatakse reegluga kaasrõhud.

Sõnas on ainult üks pearõhk ja see määratakse vastavas järjekorras:

- kui sõnas on rõhumärk, siis on pearõhk sellel silbil: **all`ergiline**
- kui on järjestikused vokaalid või vältemärk, siis on pearõhk sellel silbil: **marmelaadi, vari`an't**
- Ülejäänutel on esimene silp pearõhuline: **koera**

Edasi määratakse reegluga kaasrõhulised silbid, nii et rõhutud ja rõhulised silbid vahelduksid.

EKI ühendsõnastiku varasem seletus:

akuut märgib ebareeglipärast rõhusilpi: ekl´ektika, de´ontika, all´ergik, gramm´atiline, abrakad´abra. Reeglipärane rõhk on eesti keeles (a) esisilbil (õpik, külastatavus); (b) järgsilbil, kui selles on pikk täishäälik (meloodiline) või kui see on III vältes (ef´ektne)

EKI ühendsõnastiku uus seletus:

rõhukriips (akuut) märgib ebaharilikku rõhusilpi (silbi 1. täishääliku ees): ekl´ektika, all´ergik, gramm´atiline, abrakad´abra, anakre´ontiline, dokumental´istika. Harilikult on sõnarõhk eesti keeles esisilbil, seda eraldi ei märgita

Hääldusmärgid EKI ühendsõnastikus

EKI teatmiku järgi

Märgitakse ainult pearõhku järgsilbis ja see jäetakse märkimata, kui

- järgsilbis on pikk vokaal, nt **irooniline**, **poeetika**, **biomeetriline**
- silbil on juba III välte märk, nt **ond`atra**

„Võõrsõnades, milles rõhk on esimesel silbil, aga järgsilp on III vältes, nt **presid`ent**, esisilbi rõhku harilikult ei märgita, sest tegelikus häälduses **võib** seda tüüpi sõnades **esineda** ka rõhk järgsilbil, nt **minarett** hääldub nii **minar`ett** kui ka **m`inar`ett**.“

Kui rõhumärk on sulgudes (´), siis võib sellel silbil olla pearõhk, aga ei pea olema: **leedul(´)anna** (st et hääldatakse nii esisilbirõhuga kui ka 3. silbi rõhuga).

Rõhk Vabamorfi sõnastikus

Vabamorfi sõnastikus on ära märgitud sõnad, kus järgsilp on kolmandas vältes, aga rõhk on esisilbil.

- Järgsilbis on III välde ja rõhk on esisilbile märgitud: **1968** lemmat (v.a kohanimed, liitsõnad ja väljendid);
- Võrdluseks on sõnu, kus järgsilbis on III välde, aga rõhk ei ole esisilbile märgitud: **9294** lemmat (v.a kohanimed, liitsõnad ja väljendid);
- Lisaks on **639** sõna, kus on erandlik järgsilbirõhk, aga III völdet märgitud pole.

Vaatasime, kuidas inimhäääl ja sünteeshäääl neis gruppides rõhuga hakkama saab.

Inimhäääl vs sünteeshäääl: järgsilbirõhk III välteta sõnades

aer' oobika

all' ergiline

dil' emma



en' ergiline

fant' astiline

gal' aktika

gramm' atika



log' is'tika

medits' iiniline

nost' al'giline

peri' oodiliselt

rom' an'tiline

ros' ol'je



stat' is'tiline

Pole probleemi, rõhk esineb järgsilbil märgitud kohas nii inimhäälega kui ka sünteeshäälega loetud näidetes.

Inimhää! vs sünteeshää!:

sõnades, kus järgsilbis II vâlde ja rõhku Vabamorfi järgi esisilbil pole

Inimene hääldab esisilbirõhuga		Inimene hääldab järgsilbirõhuga
Inimhääles esisilbirõhk, aga sünteeshäälel järgsilbirõhk	Inimhääles esisilbirõhk, aga sünteeshää! rõhutab mõlemat silpi	Inimhääles ja sünteeshääles järgsilbirõhk
ekspert`iis patri`oot	laure`aat	bienn`aal diagon`aal fili`aal gladi`ool hero`iin ide`aal idi`oot koka`iin medits`iin prefekt`uur prote`iin stjuard`ess vari`an't

Inimhää! vs sünteeshää!:

rõhk III-vä!telistes sõnades, kus Vabamorfi järgi on rõhk esisilbi, sünteesisitud rõhumärgita esimesel silbil

Inimene hääldab esisilbirõhuga			Inimene hääldab järgsilbirõhuga	
Inimhääles ja sünteeshääles esisilbirõhk	Inimhääles esisilbirõhuk, aga sünteeshäälel järgsilbirõhk	Inimhääles esisilbirõhk, aga sünteeshää! rõhutab mõlemat silpi	Inimhääles ja sünteeshääles järgsilbirõhk	Inimhääles järgsilbirõhk, sünteesis rõhutab mõlemaid
advok`aat agent`uur anal`üüs indiv`iid instit`uut objekt`iiv filos`oof horis`on't horosk`oop kabin`et' katastr`oof kombin`aat komplim`en't parad`iis paras`iit patsi`ent terap`eut vundam`en't	biol`oog geol`oog labür`in't limon`aad mandar`iin marmel`aad nominat`iiv refer`aat result`aat režiss`öör sotsiol`oog šokol`aad trajekt`oor trikot`aaž	deleg`aat epis`ood pedag`oog sortim`en't	majon`ees satell`iit viktor`iin	baler`iin

Inimhäääl vs sünteeshäääl: rõhk III-vältelistes sõnades sünteesitud rõhumärgiga esimesel silbil

Inimene hääldab esisilbirõhuga			Inimene hääldab järgsilbirõhuga		
Inimhääles ja sünteeshääles esisilbirõhk	Inimhääles esisilbirõhuk, aga sünteeshäälel järgsilbirõhk	Inimhääles esisilbirõhk, aga sünteeshäääl rõhutab mõlemat silpi	Inimhääles ja sünteeshääles järgsilbirõhk	Inimhääles järgsilbirõhk, süntees esisilbirõhk	Inimhääles järgsilbirõhk, sünteesis rõhutab mõlemaid
d' eleg`aat p' edag`oog s' ortim`en't `advok`aat `indiv`iid `institut`uut `objekt`iiv f' ilos`oof h' oris`on't h' orosk`oop k' abin`et' k' atastr`oof k' ombin`aat k' omplim`en't p' arad`iis p' aras`iit p' atsi`ent t' erap`eut v' undam`en't l' abür`in't n' ominat`iiv r' efer`aat r' esult`aat r' ežiss`öör s' otsiol`oog tr' ajekt`oor	b' iol`oog g' eol`oog	`agent`uur `anal`üüs l' imon`aad m' andar`iin m' armel`aad tr' ikot`aaž `epis`ood š' okol`aad n' arkom`aan	v' iktor`iin	s' atell`iit	baler`iin m' ajon`ees

Inimhäääl vs sünteeshäääl: rõhk III-vältelistes sõnades sünteesisitud rõhumärgiga esimesel silbil

Inimene hääldab esisilbirõhuga		Inimene hääldab järgsilbirõhuga		
Inimhääles ja sünteeshääles esisilbirõhk	Inimhääles esisilbirõhuk, aga sünteeshäälel järgsilbirõhk	Inimhääles ja sünteeshääles järgsilbirõhk	Inimhääles järgsilbirõhk, süntees esisilbirõhk	Inimhääles järgsilbirõhk, sünteesis rõhutab mõlemaid
´advokaat	b´iol`oog		m´ajonees	baler`iin
´agentuur	g´eol`oog		s´atelliit	
´analüüs			v´iktoriin	
´episood				
´indiviid				
´instituut				
´objektiiv				
d´elegaat				
f´ilosoof				
h´orison't				
h´oroskoop				
k´abinet'				
k´atastroof				
k´ombinaat				
k´omplimen't				
l´abürin't				
l´imonaad				
m´andariin				
m´armelaad				
n´arkomaan				
n´ominatiiv				
p´aradiis				
p´arasiit				
p´atsient				
p´edagoog				
r´eferaat				
r´esultaat				
r´ežissöör				
s´ortimen't				
s´otsioloog				
š´okolaad				
t´erapeut				
tr´ajektor				
tr´ikotaaž				
v´undamen't				

šokolaad

- šokol`aad 

sünteshääl kolmanda välte märgiga
järgsilbil (ühendsõnastiku märgendus)

- š´okol`aad 

sünteshääl kolmanda välte märgi ja
esisilbirõhuga (Vabamorfi märgendus)

- š´okolaad 

sünteshääl esisilbirõhuga ilma 3. välte
märgita

- šokolaad 

inimhäälega

episood

- epis`ood 

süntheeshääl kolmanda välte märgiga järgsilbil (ühendsõnastiku märgendus)

- ´epis`ood 

süntheeshääl kolmanda välte märgi ja esisilbirõhuga

- ´episood 

süntheeshääl esisilbirõhuga ilma 3. välte märgita

- episood 
inimhäälega

Kokkuvõtteks

- Sõnades, kus pearõhk on nihkunud esisilbile, ei anna ühendsõnastiku hääldusmärgid seda edasi, mistõttu ei suuda ka süntesaator hääldada neid sõnu nagu inimene.
- Süntesaator ei saa hästi hakkama sõnadega, kus peaks olema nii esisilbirõhk kui ka III välde. Kuigi suudab määrata õigesti kaas- ja pearõhud, jääb kuulates mulje, et sõnas on kaks pearõhku.
- Rõhumärgi lisamine esisilbile ja vältemärgi eemaldamine nimetavas võiks olla lahendus.

Aitäh kuulamast!

Üksiksõnade süntees kuulatav siin:
elo.eki.ee/yksiksona

