

Keelekorpused keeleõppes

Jelena Kallas, PhD, vanemteadur
jelena.kallas@eki.ee

Kristina Koppel, PhD, vanemarvutileksikograaf
kristina.koppel@eki.ee

VI õpiamps
4. detsember 2025

eki.ee



Tänaſed teemad

- Mis on keelekorpus
- Millised eesti keele korpused on olemas, kust neid leida
- Korpuste kasutamine õppetöös õppematerjalide koostamisel ja harjutusvara loomisel
- Korpused Sõnaveebi ja Keeleõppija Sõnaveebi täiendava materjalina



Õpiväljund

Koolituse läbinu teab, mis on keelekorpus, kust neid leida ning kuidas neid kasutada õppematerjalide koostamisel ja harjutusvara loomisel.



Mis on keelekorpus?

” Mis on korpus?

Korpus (keelekorpus, tekstikorpus) on kirjalikest või suulistest tekstidest koosnev elektrooniline andmekogu = keele kasutusnäidete kogu. Tekst võib olla nt romaan, ajaleheartikkel, blogipostitus, tekstisõnum jm.

Sisaldab ka

- metainfot teksti kohta (autor, ilumiskoht ja -aasta jpm)
- märgenduskihte (sõna algvorm, grammatiline info, süntaktiline roll jpm)

Veebilauseid ⓘ

⚠ Veebilaused on automaatselt valitud ning võivad sisaldada vigu.

Kuigi kaamera väline disain sarnaneb eelkäijaga, on **corpuse** disainis mitmeid uuendusi.

Diplomaatilise **corpuse** vanem on suursaadik, kes on konkreetsetes riigis kõige kauem teeninud.

Korpus on küll plastist, aga peos on seade tugev ja jäik.

Remondi käigus tehti kindlaks laeva **corpuse** defektid.

Nutiseadme alumiiniumist **korpus** ei kuumene.

Puhastage lapiga ekraan ja **korpus**.

Pärgade asetamise tseremooniast võtavad osa valitsuse liikmed ja diplomaatilise **corpuse** esindajad.

Mitu korda sõna "armastus" **corpuses** esineb?

Õhuke **korpus** ja voolujooneline disain muudavad kalkulaatori käes hoidmise mugavaks.

Korpus kaitseb kaamerat põrutuste eest ning sellega võite minna isegi sukelduma.

Kaameral on veekindel **korpus**.

Pikkamisi tekkis arusaam, missuguseks võiks eesti õppijakeele **korpus** kujuneda.

Roostevabast metallist **korpus** on saadaval kahte värvi.

Telefonil on plastist **korpus**.

Sõna *korpus* kasutusnäited eesti keele veebilausete korpuses (2021)



Korpuste liigid

Kirjakeele korpused

- Eesti keele ühendkorpuste sari 2013–... (uueneb iga 2-3 aasta tagant)
 - Eesti keele veebilauseste korpus
- Vana kirjakeele korpused

Suulise kõne korpused

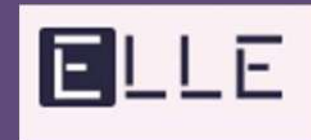
Erikorpused

- Õppekorpus
- Õppijakeele korpused
- Õpikukorpused
- Murrete korpused



Korpusanalüüsi tööriistad

Tarkvara, mis võimaldab
korpuse mitmekülgset
analüüsi.





Milleks korpusi kasutatakse (1)

Keele uurimisel, kirjeldamisel ja keelemuutuste jälgimisel:

- tähendused
- kasutussagedus
- uued sõnad ja tähendused

Loomuliku keele töötluses, nt masintõlge, juturobotid, suured keelemudelid, kõnesüntees (tekst kõneks) ja kõnetuvastus (kõne tekstiks)



Milleks korpusi kasutatakse (2)

Õppematerjalide koostamine

- harjutusvara
- näited

Iseseisev keeleõpe



Korpuspõhine keeleõpe

DDL (ingl *Data-Driven learning*)

Korpusanalüüsi tööriistade kasutamine keeleõppes

- õppijad uurivad keeleandmeid iseseisvalt
- induktiivne (õppijal lastakse näidete põhjal ise tähendusi ja reegleid tuletada)
- areneb leksikaalgrammatiline ehk sõnavara- ja grammatikapädevus
- sobib pigem edasijõudnud õppijatele
- kasulik enda kirjutatud teksti kontrollimisel
- NB! korpus peab olema vastavuses keeleoskustasemega
- eriõppeotstarbeliste korpuste vajadus



Eesti keele ühendkorpus

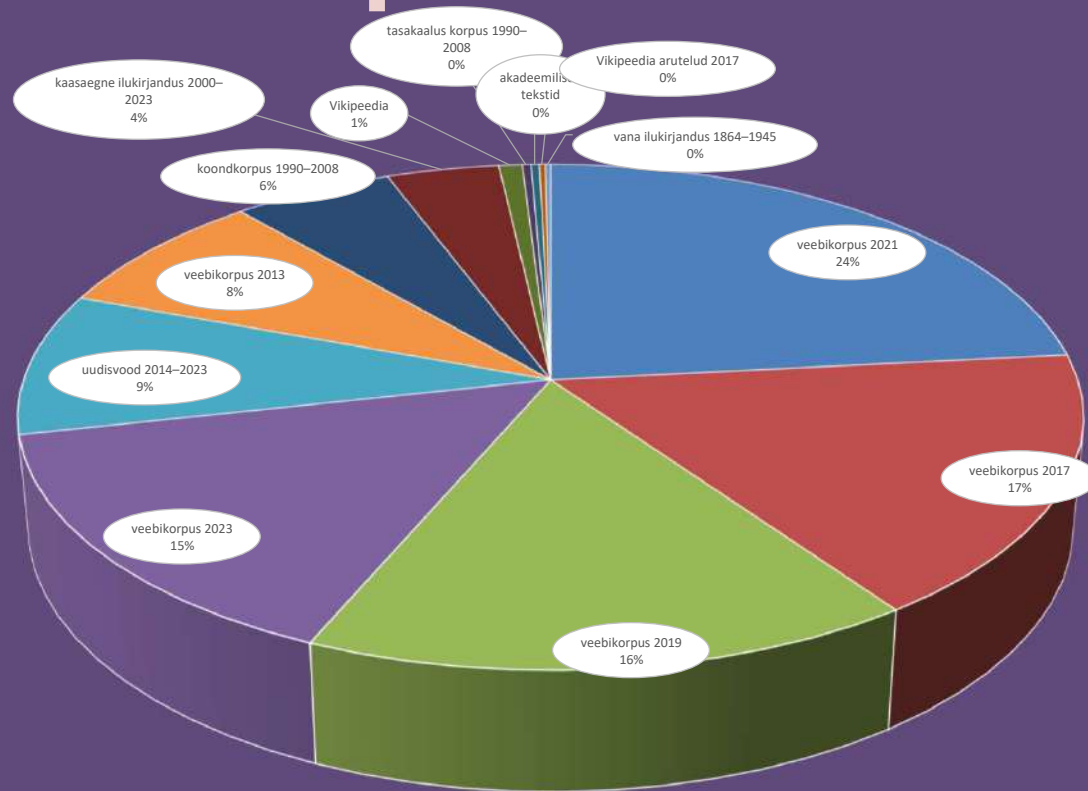
Eesti keele ühendkorpuste sari

Kaasaegseim, mahukaim ja žanriliselt mitmekesiseim eestikeelsete digitekstide kogu. Valminud EKI ja UK/Tšehhi tarkvarafirma Lexical Computing Ltd. koostöös.

	Tekstisõnu	Lauseid	Lõike	Dokumente
eesti keele ühendkorpus 2023	3,8 mld	255 mln	83 mln	15 mln
eesti keele ühendkorpus 2021	2,9 mld	197 mln	64 mln	12 mln
eesti keele ühendkorpus 2019	1,8 mld	120 mln	35 mln	6 mln
eesti keele ühendkorpus 2017	1,3 mld	88 mln	27 mln	3 mln
eesti keele ühendkorpus 2013	563 mln	38 mln	7,5 mln	700 tuh

”Eesti keele ühendkorpus 2023

Allkorpus	Tekstisõnu
veebikorpus 2021	884 524 889
veebikorpus 2017	638 470 425
veebikorpus 2019	613 775 513
veebikorpus 2023	571 221 646
uudisvood 2014–2023	344 463 562
veebikorpus 2013	302 950 928
TÜ koondkorpus 1990–2008	212 424 035
kaasaegne ilukirjandus 2000–2023	148 559 737
Vikipeedia	30 747 204
TÜ tasakaalus korpus 1990–2008	11 631 801
akadeemilised tekstid	11 026 315
Vikipeedia arutelud 2017	75 71 584
vana ilukirjandus 1864–1945	7 421 569





Mis tekstid veebikorpuses on?

Meediaportalide ja perioodikaväljaannete tekstid

- Delfi, Postimees ja Õhtuleht, ERRi portaal, Saarte Hää, Eesti Ekspress, Maaleht, Äripäev, Õpetajate Leht, Sirp, Horisont, Tehnikamaailm, Eesti Loodus, Pealinn, Kesknädal, Lääne Elu, Eesti Kirik, Hiiu Leht, Põhjarannik

Ametlikud ja haldustekstid

- Riigi Teataja, Euroopa Parlament, Riigikohus, Riigikogu, Vabariigi Valitsus, Riigikontroll, Õiguskantsleri Kantselei, Välisministeerium, Rahandusministeerium, Siseministeerium, Maksu- ja Tolliamet, Politsei- ja Piirivalveamet, Statistikaamet, Töötukassa, Terviseamet, Transpordiamet, Riigi Infosüsteemi Amet, Eesti Keele Instituut

Asutuste kodulehed ja veebipoed

- Tartu Ülikool, Tallinna Tehnikaülikool, Kirjandusmuuseum, Eesti Kunsti Muuseum, Vanemuine, Linnateater, TV3, Apollo, LHV, Kaup24, Photopoint, Tradehouse

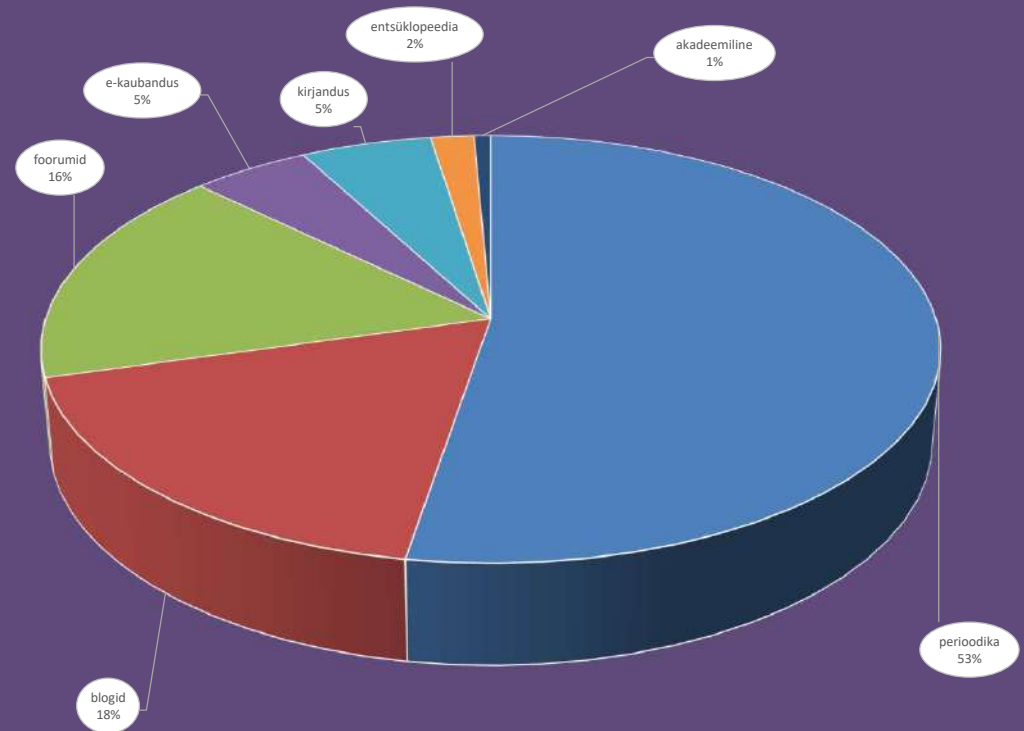
Blogid ja foorumid

Teaduskeel

- Keel ja Kirjandus, Eesti ja soome-ugri keeleteaduse ajakiri, Folklore, Eesti Arst, Eesti Haridusteaduste Ajakiri, Ajalooline Ajakiri, Mäetagused, Eesti Majanduspoliitilised Väitlused, Agraarteadus

Žanrid

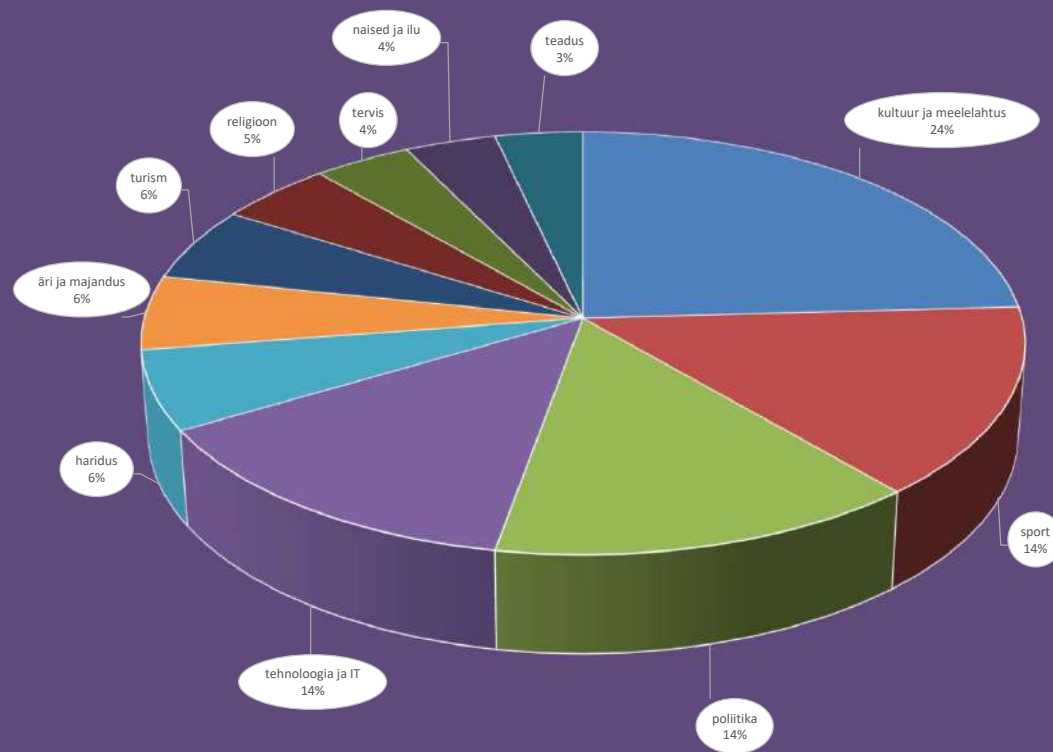
Žanr	Tekstisõnu
perioodika	1 574 063 873
blogid	537 993 698
foorumid	497 219 390
kirjandus	161 255 917
e-kaubandus	145 984 966
entsüklopeedia	53 420 451
akadeemiline	20 329 641



Teemad

Teema	Tekstisõnu
kultuur ja meelelahtus	382 851 434
sport	231 393 259
poliitika	224 094 079
tehnoloogia ja IT	217 117 346
haridus	96 707 296
äri ja majandus	89 853 689
turism	89 325 169
religioon	75 187 695
tervis	63 852 732
naised ja ilu	59 374 774
teadus	57 064 418

Lisaks: ehitus ja kinnisvara, loodus ja keskkond, juura, seks, loomad, pere ja kodu, ajalugu, mängud, toit ja jook





Kirjanduse allkorpused



Vana ilukirjandus 1864–1945 (7 421 569 tekstisõna)

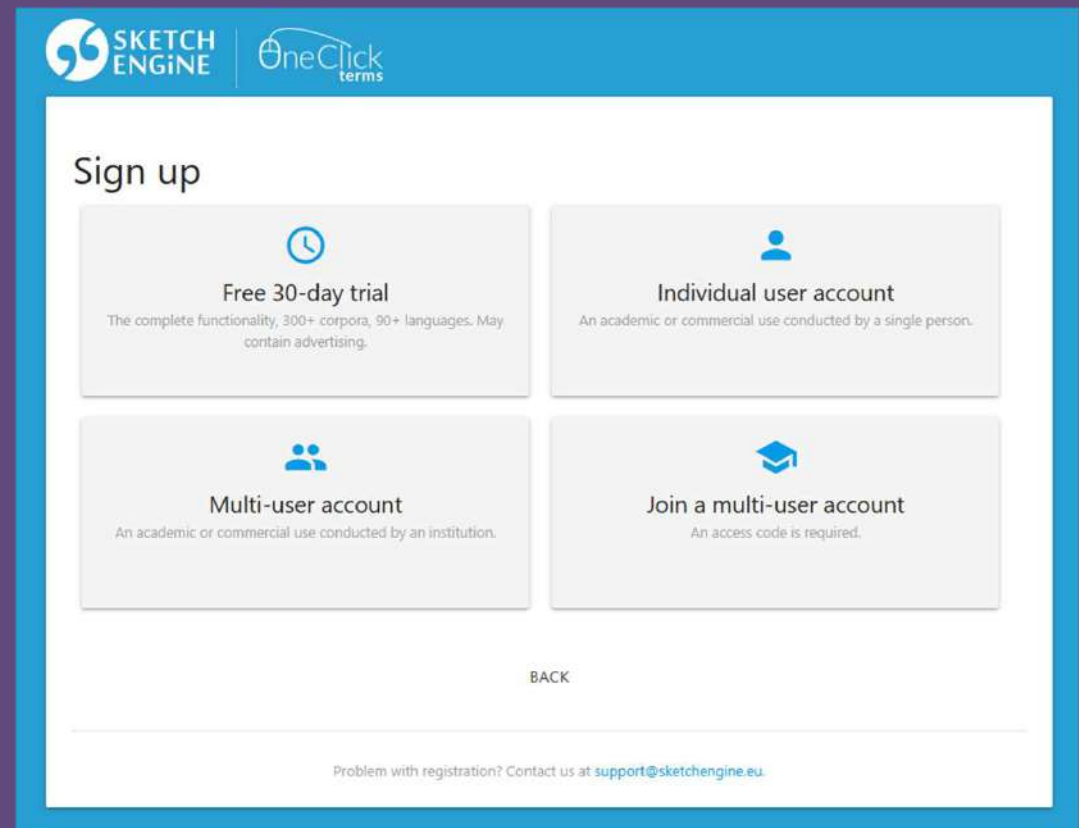
- **Autorid:** Eduard Bornhöhe, Eduard Vilde, Elisabeth Aspe, August Kitzberg, Juhan Liiv, Carl Robert Jakobson, Anton Hansen Tammsaare, Friedrich Robert Faehlmann jpt

Kaasaegne ilukirjandus 2000–2023 (147 273 709 tekstisõna)

- **Autorid (originaalid):** Tõnu Õnnepalu, Dagmar Lamp, Leelo Tungal, Karl Ristikivi, Indrek Hargla, Andres Ehin, Jaan Undusk, Aimée Beekman, Andrus Kivirähk, Raimond Kaugver jpt
- **Autorid (tõlked):** Emily Brontë, Rosamunde Pilcher, Rex Stout, Jane Austen, Agatha Christie, Virginia Woolf, Umberto Eco, William Shakespeare, Tove Jansson, Stephen King, J. K. Rowling jpt



Eesti keele ühendkorpuste sarja saab analüüsida korpusanalüüsi tööriistaga Sketch Engine (sketchengine.eu).



Näidete allikas

CONCORDANCE Estonian National Corpus 2023 (Estonian NC 2023)

simple keelekorpus • 436
0.12 per million tokens • 0.000012%

Details Left context KWIC Right context

1	Web 2013 • peri... alub üle keelemudelis oleva. Statistilist keelemudelit treenitakse väga suurte	keelekorpuste	põhjal (vähemalt paar miljonit sõna, soovitavalt palju rohkem). Korpuste aluse
2	Web 2021 • peri... t, autorsuse tuvastamine, tekstide sisukokkuvõtete tegemine, masintõlge jne.	Keelekorpuse	alusanalüüs ehk vormistiku- ja sõnaliigimärgendus ei ole aga siiani lahendatu
3	Web 2021 • peri... õnaliigimärgised on ELANis ühendatud vormi- ja lauseehituse infoga. Suurte	keelekorpuste	töötlemise iseärasused mõjutavad sõnaraamatutegijate tööd, kellele on olulin
4	Web 2021 • ===N... olemasolust, kasvatada nähtavust, olla noortele kasutatav ja atraktiivne. Ka	keelekorpuse	loomisel on see hetkel parim allikas. Arutati keeleterminoloogia teemadel – ku
5	Timestamped 201... nil oleks rahvusliku ärkamisaja lõpul olnud kasutada selline võimas relv nagu	keelekorpus	, küllap nad oleksid oma aja keelevaidlusi pidades (et kas hää ja pää või hea
6	Timestamped 201... ja pea) teineteist sellega nüpeldanud ajalehtede juhtkirjades ja pealugudes.	Keelekorpuse	jumaldamine tõe allikana on progressiuseks liberaalid viinud järeltuseneni, et l
7	Timestamped 201... jumaldamine tõe allikana on progressiuseks liberaalid viinud järeltuseneni, et	keelekorpus	on sisepõlemismootoriga sõiduk ja sõnastik veohobune, kelle tõu revolutsioori
8	Timestamped 201... i. Aktuaalse näite varal hakkab olema nii, et kuna suur osa Eesti elanikest (ja	keelekorpuse	täienemisesse panustavad iga päev kõik, kel sõrmed ja arvuti) hääldab ja kirj
9	Timestamped 201... ametiasutustes kui ka meedias ja kultuurisfääris. " Kuidas seda küll teha, kui	keelekorpuses	tuulavad masinad on pandud tootma argumente kõige lubatavuse kasuks, ko
10	Web 2023 • acad... i. Vikipeedia on üks suurimaid tekstide kogumeid eesti keeles. Seda kasutab	keelekorpused	arvutilingvistika teadus ja see on aluseks paljudele statistilistele andmetele. V
11	Web 2023 • ===N... nõuannet, koolitatakse ametiasutuste töötajaid. Keeletehnoloogiline areng ja	keelekorpused	on lisanud keeleuurimise võimalustesse uue dimensiooni. Peamisteks proble
12	Web 2023 • ===N... atakse eesti keele sõnaraamatuid, ülevaateid ja käsiraamatuid, täiendatakse	keelekorpuseid	ja andmebaase, tellitakse valdkonna uuringuid. arendatakse keelehooldesüst
13	Academic Texts ... tibe ja mõru leksikaalsed suhted sügavamalt ja laiemat analüüsi, hõlmates ka	keelekorpuseid	, lisaküsitlusi jms. Seda, et sama tähendusvälja eri aspektide väljendamiseks
14	Academic Texts ... luse põhistamiseks vääraks teema põhjalikku uurimist, millesse võiks kaasata	keelekorpused	, lisaküsitlused jne. Nagu tulemustest selgus, siis viiendaks põhimaitesõnaki
15	Academic Texts ... eelist käitumist (Cameron 1995: 48). Põhjus on lihtne: sõnaraamatus teeme	keelekorpuse	põhjal suure hulga inimeste keelekasutusest, mitte ei põhine üksikisikute arva
16	Academic Texts ... põhimõtte keelekorraldusele on enne soovitusi andmist uurida sõna kasutust	keelekorpuse	põhjal ja analüüsida, kas tegelikult keelekasutuses on keelelendi mõistmisega j

Naabersõnade otsing

WORD SKETCH Estonian National Corpus 2023 (Estonian NC 2023)

tehisaru as common noun 1,004× Sorted by frequency X

Constructions

omadussõnaga	subjektina	"tehisaru" omastavas	ja/või
generatiivne 9 9.3 ... generatiivse tehisaru	suutma 7 0.2 ... Tehisaru suudab	tehisaru_kasutamine 34 11.5 ... tehisaru	tehnoloogia 4 1.3 ...
võimas 7 1.9 ... arusaam , et võimsate tehisarude arendamist tuleks hakata	asendama 3 2.2 ...	tehisaru_abi 13 10.2 ... tehisaru abiga	juturobot 3 9.2 ...
		tehisaru_arendamine 12 10.1 ... tehisaru	masintõlge 3 8.3 ...
		tehisaru_mudel 9 9.7 ... tehisaru mudelite	kratt 3 6.7 ...
		tehisaru_programm 9 9.7 ... tehisaru	küberturvalisus 3 5.5 ...
		tehisaru_loomine 9 9.7 ... tehisaru loomisel	robot 3 4.1 ...
		tehisaru_mõju 8 9.5 ... tehisaru	internet 3 1.7 ...
		tehisaru_teema 7 9.3 ... tehisaru teemal	
		tehisaru_tulek 7 9.3 ... et tehisaru tulek	
		tehisaru_arendaja 6 9.1 ...	
		tehisaru_ajastu 6 9.1 ...	
		tehisaru_areng 6 9.1 ...	

Trendid

Võimaldab analüüsida sõna kasutussageduse muutumist ajas. Kasutuse suurenedes on trend positiivne (**roheline nool** suunaga üles) ning kasutuse langedes negatiivne (**punane nool** suunaga alla)

TRENDS			Estonian Trends			
	Lemma	Trend ↓	Frequency			
1	tehisaru	↗	9,556	...		
2	prompt	↗	81	...		
3	nakatumisjuht	↘	128	...		
4	fotoreis	↘	170	...		
5	drooniüksus	↗	136	...		
6	kaalulangetusravim	↗	327	...		
7	putini	↗	696	...		
8	sergei	↗	255	...		
9	regionaal	↗	2,010	...		
10	raamatublogi	↘	117	...		
11	nakatumisrisk	↘	198	...		
12	karantiininõue	↘	248	...		
13	putin	↗	1,681	...		
14	aleksandr	↗	113	...		
15	jäätmereform	↗	393	...		
16	cookies	↘	113	...		
17	haigekassa	↘	6,378	...		
18	maaeluminister	↘	2,068	...		
19	koroonahaige	↘	1,850	...		
20	testimispunkt	↘	536	...		
21	viiruselevik	↘	262	...		
22	koroonates	↘	83	...		
23	päisepilt	↗	104	...		
24	generatiivne	↗	485	...		
25	droonivastane	↗	170	...		
26	droonioperaator	↗	272	...		
27	ühispartei	↗	472	...		
28	kaudtuleüksus	↗	504	...		



Eesti keele veebilausete korpus

Eesti keele veebilausete korpus on väljavõte eesti keele ühendkorpusest 2023, mida kasutatakse:

- Sõnaveebis veebilausete kuvamiseks
- SkELLis

Sisaldab eesti keele ühendkorpusest spetsiaalse tööriista abil valitud lauseid, mis vastavad teatud tingimustele (lause/sõnade pikkus, sõnade sagedus korpuses, teatud sõnaliigi kordumine lauses jms).



Korpused, millele täna keskendume

- **Õppekorpused**, eelkõige Eesti keele veebilausete korpus
kus: SkELL, Sõnaveeb
kellele: õpetajad, õppijad B1–C1
- **Õpikukorpused**
kus: KORP, Keeleõppija Sõnaveeb
kellele: õpetajad, õppijad A1–C1
- **Õppijakeele korpused**
kus: KORP, ELLE
kellele: pigem õpetajad
- **Kirjakeele korpused**
kus: KORP, Keeleveeb, Sketch Engine
kellele: õpetajad, õppijad al B2



SkELL

(Sketch Engine for Language Learning)

www.skell.sketchengine.eu



Näidisharjutus 1. Tähenduste otsimine (*Examples*)

Tahate pärast tänast koolitust oma õpilastele tutvustada korpuse kasutamist SkELLis. Õpilaste jaoks on *korpus* selles tähenduses uus sõna. Te palute õpilastel SkELLis analüüsida lauseid, kus sõna *korpus* esineb, ja tuletada sõna tähendusi.



www.skell.sketchengine.eu



Võrdlus EKI ühendsõnastikuga

Võrrelge tähendusi EKI ühendsõnastiku tähendusjaotusega Sõnaveebis.

Kas Sinu leitud tähendused on olemas ka EKI ühendsõnastikus?

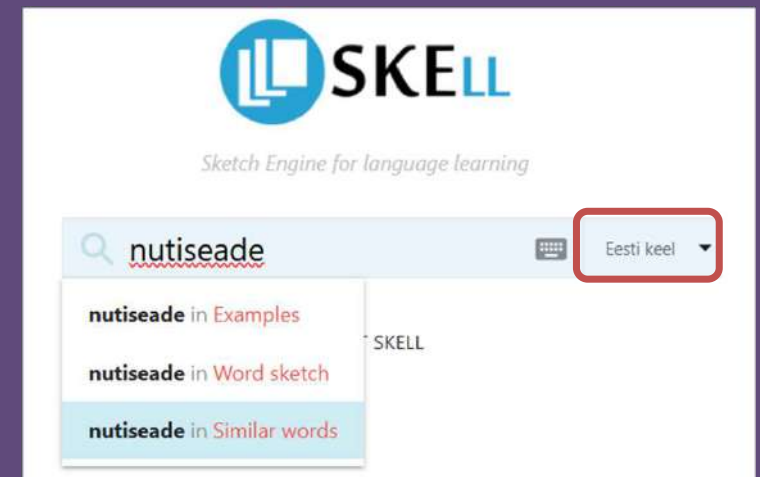
Kas EKI ühendsõnastikus on veel tähendusi, mida SkELLis ei olnud?

” Näidisharjutus 2. Sõnapilv (*Similar words*)

Teie tunni teema on **nutiseadmed**.

Tunni alguses suunake õppijad SkELLI sõnapilve juurde ja paluge neil paaris arutada, kuidas need sõnad võiks omavahel seotud olla.

Kui tegemist on õppijate jaoks tundmatute sõnadega, suunake nad EKI ühendsõnastikku (sonaveeb.ee).



www.skell.sketchengine.eu



Näidisharjutus 3. Näitelausete otsimine harjutuse (*Examples*)

Teie tunni teema on püsiühendid ning te otsite harjutusvara loomisel lauseid, milles on kasutatud teatud püsiühendit (nt *hakkama saama*) konkreetsetes vormis (nt *saadi hakkama + millega*).



Tärniga otsing Sõnaveebis ja SkELLis

Teie tunni teema on digimaailm. Te tahate õpilastega otsida e-ga algavaid sõnu, nt *e-pood*, *e-valimised*. Õpilase ülesanne on leida vähemalt viis e-ga algavat sõna ja koostada neist üks sidus tekst.

(otsing: e-*)



Näidisharjutus 4. Sõnakasutus kontekstis

Tunni teema on püsiühendid ja õppija ülesanne on leida püsiühendeid sõnaga *silm*.

Suunake ta SkELLi näitelausete juurde, kus ta peab läbi lugema 20 näitelauset, mille ta on otsinud päringuga **.*silm** ning leida kolm ühendit või naabersõna, mis on tema jaoks uued.

Tundmatute sõnade korral võib kasutada EKI ühendsõnastikku Sõnaveebis.

The screenshot shows the search results for the query **.*silm** on the EKI Sõnaveeb website. The search bar at the top shows the query **.*silm** and the language is set to Estonian. Below the search bar, there are tabs for 'Examples', 'Word sketch', and 'Similar words'. The main content area displays a list of 21 example sentences, each with the word **silm** or its derivatives highlighted in red. The sentences are numbered 1 through 21. The list includes various contexts where the word is used, such as 'Linnaelanikule on oluline, et linna juhitakse ausalt ja hästi, ühishuvisid **silmas** pidades.', 'Hoia vaid **silmad** lahti, viidad ja sildid juhatavad sind kohale.', 'Elu vanglas on autor jälginud eeskätt oma noore peategelase **silmade** kaudu.', 'Ta pigistab **silmad** hästi kõvasti kinni ja teeskleb magamist.', 'Ripsmete koolutamine teeb **silma** visuaalselt suuremaks.', 'Tüdruk raputas energiliselt pead, nii et juuksesalgud ümber **silme** lendasid.', 'Kahe tunniga valmis tehtud filmilõiku saab iga noor režissöör pärast oma **silmaga** üle vaadata.', 'Naine sulges **silmad** ja ta köht liikus hingamise taktis vaevumärgatavalt.', 'Tänu sädelevatele **silmadele** paistad kindlasti teise pidutsejate seast silma.', 'Tänu sädelevatele silmadele paistad kindlasti teise pidutsejate seast **silma** .', 'Reklaamiseaduse eelnõu vaadeldes torkab **silma** suur hulk ebatäpsusi.', 'Eesti on püüdnud maailmas **silma** paista innovaatilise maksusüsteemi ja liberaalse väliskaubandusega.', 'Lapse **silmad** läksid tuimestuse peale tagurpidi – tunne oli õõvastav.', 'Kui **silma** valulikkus ja ärritus ei kao, pöörduge silmaarsti poole!', 'Teismelisena olin hirmus arg, püüdsin võimalikult vähe **silma** torgata.', 'Valgusti peab kuvari suhtes paiknema nii, et see ei pimestaks **silmi** .', 'Ka valgustatud tavalisel torkab helkuriga jalakäija palju paremini **silma** .', 'Parki korda tehes pidasime **silmas** eelkõige laste huve.', 'Tööpäeva lõpuks on **silmad** alatasa väsinud ja ärritunud.', 'Euroopa linnades ringi käies torkab **silma** üks kindel tendents.', 'Autori definitsiooni kohaselt on intervjuu andmete kogumine, mis toimub **silmast-silma** olukorras.'

Naabersõnad EKI ühendsõnastikus

Naabersõnad ehk
kollokatsioonid on sageli koos
esinevad sõnad, nt *pisike
päkapikk, päkapikke uskuma*.

- 10 000 sõna juures
- reastatud sageduse järgi
- rühmitatud sõnaliikide kaupa
- illustreerivad näitelaused

et **päkapikk**  nimisõna

EKI ühendsõnastik 2025

et B2 punase mütsiga väike mees, kes jõulude ajal lastele kingitusi toob, jõuluvana abiline

Näited
Päkapikk tõi sussi sisse kommi. 



Naabersõnad

omadussõnaga
väike **päkapikk** | pisike | punane | tõeline | tore | tubli ...

nimisõnaga
jõuluvana ja **päkapikud** ...

tegusõnaga
päkapikud käivad | toovad | tulevad | jagavad ...
päkapikke uskuma | ootama ...

^ Näita vähem

<https://sonaveeb.ee/search/lite/est/p%C3%A4kapikk/>



Näidisharjutus 5. Naabersõnad SkELLis (*Word Sketch*)

Õppija ülesanne on kirjutada jõuluteemaline tekst, kus peategelased on päkapikud. Suunake ta SkELLI naabersõnade (*Word Sketch*) juurde ning paluge tal vaadata, milliste tegusõnade ja omadussõnadega *päkapikk* koos esineb.

The screenshot shows the SkELL web application interface. At the top, there is a search bar with the word 'päkapikk' entered. Below the search bar, there are tabs for 'Examples', 'Word sketch', and 'Similar words'. The 'Word sketch' tab is selected. Below the tabs, there is a section titled 'päkapikk' with a dropdown menu set to 'nimisõna' and a 'Show context' checkbox. The main content area is divided into two columns: 'omadussõnad' (adjectives) and 'käändumatud omadussõnad' (non-inflecting adjectives). The 'omadussõnad' column lists 15 items, each with a number, a word, and a description. The 'käändumatud omadussõnad' column lists 3 items, each with a number, a word, and a description.

omadussõnad		käändumatud omadussõnad		
1.	usin	usinad päkapikud	1.	argentiina
2.	piiluv	piiluvate päkapikkudega	2.	võrumaa
3.	hiiliv	hiilivad päkapikud	3.	hiina
4.	ninatark			
5.	tillukene			
6.	virk			
7.	usinam			
8.	lahke	lahkeid päkapikke		
9.	töökas	töökad päkapikud		
10.	lustlik	mängutoa lahti töönud lustlikud päkapikud		
11.	hoolas	hoolsad päkapikud		
12.	sinimustvalge	100 sinimustvalget päkapikku		
13.	toimetav			
14.	väsimatu			
15.	pükniline			

Näidisharjutus 6. Naabersõnad (*Word Sketch*)

Teie tunni teema on inimeste tunded. Vaadake koos õpilastega SkELList sõna ***armastama*** naabersõnu (*Word Sketch*). Keda või mida kõige rohkem armastatakse?



Sõnaveebiga lingitud korpused

www.sonaveeb.ee

Eesti veebilausete korpus Sõnaveebis

Lisamaterjal leksikograafide lisatud kasutusnäidete kõrval.

Kasulikud eriti sõnade puhul, millele leksikograaf pole näidet lisanud (EKI ühendsõnastikus selliseid üle 40 000, nt *šamaan*).

Veebilaused

⚠ Veebilaused on automaatselt valitud ning võivad sisaldada vigu.

Kitsamas mõistes tähendab ekstaas **šamaani** hinge ajutist siirdumist teispoolsusse.

Joonised tegid transiseisundis alateadvuse sügavates soppides ekslevad **šamaanid**.

Šamaani sotsiaalne positsioon on erinevates kultuurides olnud erinev.

Lõppev aasta on **šamaani** sõnul olnud eestlastele raske.

Kontserdist jäi meelde **šamaani** raju tants ümber tule.

Valged **šamaanid** olid taevase maailma preestrid.

Õigel **šamaanil** on kogu aeg õpilasi.

Peale lühikest und rääkisime **šamaaniga** kogetust.

Tuli meelde **šamaani** soovitus olla uudishimulik.

Laulu ajal kuuleb **šamaan** vastuseid küsimustele.

Kalju lähedal pühal väljal põletati ja maeti **šamaane**.

Šamaanid on vahendajad inimeste ja vaimude maailma vahel.

Ekstaas kitsamas mõistes tähendab **šamaani** hinge ajutist siirdumist teispoolsusse.

Näidisharjutus 7. Veebilaused kasutusnäidetenä

Õpilane vaatab Sõnaveebist sõna *kodukino*, aga sellel pole näitelauseid. Suunake ta veebilausete rubriiki.

The screenshot shows the EKI online dictionary interface for the word "kodukino". The main content area displays the definition: "kodune audio-videoaparatuur, mis võimaldab vaadata filme kino helisüsteemile lähedase kõlaga" (home audio-video equipment that allows watching films with a sound close to cinema sound). The right sidebar contains several sections: "Sõnavormid" (Word forms) showing "kodukino" and "kodukinosid"; "Päritolu andmed" (Origin information) showing "et: kodukino" and "en: home cinema"; "Sõna seosed" (Word relations) showing "Liitsõna järelosa (põhisõna): kino" and "Liitsõna esiosa (laiend): kodu"; and "Veebilaused" (Web examples) which is highlighted with a red box. The "Veebilaused" section contains a warning icon and text: "Veebilaused on automaatselt valitud ning võivad sisaldada vigu. Kui **kodukino** sai lõpuks püsti, siis tagumised kõlarid ei..." (Web examples are automatically selected and may contain errors. If **kodukino** got up, then the rear speakers don't...).

Näidisharjutus 8. Näitelaused võõrkeelsetele vastetele

Õpilane otsib Sõnaveebist ingliskeelset sõna *butterfly*, mille sõnaartiklis pole näitelauseid ega naabersõnu. Suunake ta veebilausete rubriiki ning paluge tal vajutada SkELLi logole ja leida sõnaga *butterfly* kolm lauset.

Veebilaused

 Veebilehelt SkELL saad vaadata sõna kasutusnäiteid, naabersõnu ja sarnase tähendusega sõnu. Need on automaatselt valitud ning võivad sisaldada vigu.





Õpikute korpus Keeleõppija Sõnaveebis

- Eesti keele A1–C1 õpikute lausete korpus 2021
- Sisaldab A1–C1 keeleoskustasemega märgendatud eesti keele täiskasvanud õppijatele suunatud õpikute tekstidest eraldatud terviklikke lauseid



Õpikukorpused EKI KORP

korp.eki.ee

Hetkel ka aadressil korp.keeleressursid.ee

Näidisharjutus 9. B1-taseme näitelaused

Räägite tunnis eesti keele homonüümidest. Tahate koostada harjutust, kus oleks segamini laused, milles esinevad erinevad homonüümid (nt *maks*, *tamm*, *hall*). Kasutage KORPi, valige sealt **õpikukorpused** ja otsige B1-taseme lauseid, milles esineb sõna *hall*.

Tänapäevased tekstid | Ajaloolised | Ühestatud | Puudepangad | Õpikukorpused | Keeleõppe korpused | veel ▾ Vastukaja Viita Korpil

  3 korpust valitud (kõik) — 819.63K üksust 819.63K sõne hulgast

Lihtne Laiendatud Edasijõudnuile Võrdle

algvorm ▼ on ▼
hall Aa
või

keeleoskustase ▼ on ▼
B1 ▼
või



Õppijakeele korpused: EKI KORP

EKI õppijakeele korpused

EMMA korpuse EKI alaosa (kuni 2022 kogutud tekstid)

kus: EKI KORP

suurus: 10747 L2 teksti, sh tasemetööd (4. ja 7. klass) ja eksamitööd (9. ja 12. klass) ja 1198 L1 teksti

sisu: pseudonüümitud ja morfosüntaktiliselt märgendatud

EKI K2 keeleõppija veamärgendatud õppijakeele korpus

kus: EKI KORP

suurus: 1503 lauset

EKI K2 keeleõppija paralleellausete korpus

kus: EKI KORP

sisu: võrdlevalt originaallaused ja korrigeeritud laused

suurus: 1503 lauset

Projekti “Professionaalne eestikeelne õpetaja mitmekeelses klassis” seireks kogutud õppijakeele andmed

Uued materjalid (tulekul)

EKI noore ja täiskasvanud keeleõppija K2 korpus

kus: EKI KORP

sisu: 2023. ja 2024. aasta noote keeleõppija tasemetööd (4. ja 7. klass) ja eksamitööd (9. ja 12.klass)

täiskasvanud keeleõppijate tasemetööd (A2–C1), igalt tasemelt 200 tööd

Veamärgendatud korpuse kasutamise võimalused

Otsing
veatüübi järgi
teigusõna +
ma-infinitiiv

The screenshot shows the Korpusprõngusüsteem (Korpus Link System) interface. The search filters are set to 'VTAMARKEERID' and 'TEGUSÕNA TUNNUS'. The search results display a list of words with their corresponding grammatical forms and frequencies. The interface is in Estonian and includes a search bar, filters, and a results table.

Veatüüp	Teigusõna	ma-infinitiiv
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid
vtamareerid	vtamareerid	vtamareerid

EKI veamärgendatud õppijakeele korpus

R:VERB:FORM + category MA-infinitive

E2 ÕPPIJATE VEAMÄRGENDATUD TEKSTIDE KORPUS (ei b)

Selleks peab oskama	joonistama	ja kirjutama.
Selleks peab oskama	joonistama	ja kirjutama .
Ajalehe teema on " Mida ei tohi	tegema	koolis ".
me, me võime ka vaatamisväärsused	vaatama	.
Ma tahan sind	kohtuma	koos minu sõbradega ja me võime h
u sõbradega ja me võime hästi aega	veetma	.
Me saame võrkpalli mängida,	sõitma	jalgrattaga ja veel midagi.
ina aga unustas 1 asi, et saame ikka	kohtuma	ja teha midagi koos!
ad sa[3] oskama fottoid võta ja neid	lõikama	.
ii sa tahad, võime minu sõbraga kaa	kohtuma	, Sa võid minu kojus puhkata ja süüa
võime ka arvutimangu ja palli mängu	mängima	.
Oli väga tore vaadata ja pilti	tegema	!

WORD ATTRIBUTES

WORD: joonistama

WORD ID: 4

NUMBER: _

LEMMA: joonistama

ERROR CORRECTION SUGGESTIVE: false

ANNOTATOR ID: 1

ERROR CORRECTION: joonistada

ERROR TAG: R:VERB:FORM

DEPENDENCY RELATION: xcomp

DEPENDENCY HEAD (ID): 3

VERB MARKER: ma

CASE: _



Õppijakeele korpused: ELLE (TLÜ)



ELLE õppijakeele tekstikogud

ELLE on Tallinna Ülikoolis arendatav keskkond eesti keele õpetamiseks, õppimiseks ja keeleanalüüsiks.

Keskkonna aluseks on **eesti vahekeele ehk õppijakeele korpus EVKK**, mis on eesti keele kui teise keele ja võõrkeele õppijate kirjalike tekstide kogu.

ELLE-s saab kirjutatud ja analüüsitud tekste EVKK-le loovutada.

ELLE õppijakeele tekstikogud

Mine veebilehele

<https://elle.tlu.ee/tools/wordlist>

Rubriigist "**Tekstid ja tööriistad**" valige funktsioon "**Sõna kontekstis**" ja koostage eesti keele kui teise keele riiklike tasemetööde B1-taseme tekstide alusel sõnade sagedusloend.

The screenshot shows the ELLE website's word list tool. The top navigation bar includes links for 'Tekstihindaja', 'Tekstid & tööriistad' (highlighted), 'Lingikogu', and 'Keskkonnast'. Below this, a breadcrumb trail shows 'Tekstid & tööriistad' > 'Sõnaloend'. The main interface features a search bar labeled 'OTSI TEKSTE' and a tabbed menu with 'Sõnaloend' (selected), 'Sõna kontekstis', 'Naabersõnad', 'Sõnaanalüsaator', and 'Mustreidja'. The 'Sõnaloend' tab is active, displaying three columns: 'Alamkorpus', 'Teksti andmed', and 'Autori andmed'. The 'Alamkorpus' column has checkboxes for 'kõik', 'K2 riiklikud eksamitööd' (checked), 'K2 olümpiaditööd', 'K2 keeleõpe', 'K1 eesti keel', and 'K1 vene keel'. The 'Teksti andmed' column has dropdowns for 'Liik' (set to 'K2 riiklik eksamitöö'), 'Keel' (set to 'eesti'), and 'Tase' (set to 'B1'). The 'Autori andmed' column has dropdowns for 'Vanus', 'Sugu', and 'Haridus'.



EKI Õpetaja tööriistad



Õpiku- ja õppijakeele- korpuste näited keeleoskustasemete grammatikapädevuse kirjeldustes

Eesti Keele Instituut ÕPETAJA TÖÖRIISTAD		Sõnavara	Grammatika	Kasutusolukorrad	Õppeteksti hindamine	Õppevara	Projektist	Sõnaveeb Eesti Keele Instituut
Keeletase:		B1						
Peakategooria:		VORMIMOODUSTUS						
Alamkategooria:		TEGUSÕNA						
Teema:		kindla kõneviisi olevik jaatavas kõnes						
Oskab-lause:		Oskab tuttavatest tegusõnadest moodustada kindla kõneviisi oleviku jaatava kõne vorme (nt <i>lahkub, arvate, hakkavad</i>), et väljendada kõnehetkel ja tulevikus toimuvat tegevust. Tegusõnade <i>meie</i> -vorme (nt <i>lähme, hakkame</i>) oskab kasutada ka ennast kaasava ettepaneku tegemisel. Võib eksida vormide moodustamisel, eriti astmevahelduslikes tegusõnades.						
Tunnused:		kõneviis kindel tegumood isikuline aeg olevik pööre 1,2,3 arv ainsus,mitmus kõneliik jaatav						
Süntaktiline funktsioon:		Õeldis						
Allikas		Näitelaused						
projekti grammatikaekspertid		Ma käin toidupoes peaaegu iga päev. Järgmisel suvel sõidame perega Saksamaale. Hakkame juba minema!						
Kooliõpikute lausete korpus 2020		Marju lõpetab kirja, sest peab koju minema. Hoian teie lapsi Tartus Maarjamõisas. Päike tõuseb jaanipäeval kell 4.05.						
Eesti vahekeele korpus		<i>Algupärane lause</i> Ma kasutan interneti harva aga ikka kasutan, et saada teada, mis maailmas toimub. Kodus ma loen erinevaid raamatuid. Praegu ta töötab perearstikeskuses, ta on perearst.			<i>Parandatud lause</i> Ma kasutan internetti harva, aga ikka kasutan, et saada teada, mis maailmas toimub.			



Keeleveeb

www.keeleveeb.ee



Näidisharjutus 10. Näited Keeleveebist

Mine veebilehele www.keeleveeb.ee.

Otsi korpuse rubriigist allkorpus Kroonika näiteid sõnaga *vinge*. Milliseid nimisõnu selle omadussõnaga kirjeldatakse?

KEELEVEEB

Otsi sõna: Eesti

Sõnastikud Kroonika

☒ Korpused

- ☐ Tasakaalus korpus
- ☐ etTenTen
- ☐ Ilukirjandus
- ☐ Jututoad
- ☐ Foorumid
- ☐ Uudisgrupid
- ☐ Kommentaarid
- ☐ Eesti Päevaleht
- ☐ Postimees
- ☐ (SL) Õhtuleht
- ☐ Eesti Ekspress
- ☐ Maaleht
- ☐ Valgamaalane
- ☐ Luup
- ☐ Lääne Elu
- ☒ Kroonika
- ☐ Horisont
- ☐ Eesti Arst
- ☐ Arvutustehnika ja Andmetöötlus



Kokkuvõtteks

- Korpused sisaldavad keeleandmeid, millele toetudes on loodud eri tüüpi keeleressursse (sh sõnastikud)
- Eri tüüpi korpused pakuvad eri tüüpi infot
- Korpustest võiks saada õpetaja abiline
- Korpuseid ei pea kartma – julgustame ka õppijaid korpusi kasutama

“

AITÄH!

”