Liivike. Kuidas kasutada

(25.11.2014)

Programmi fail on **liivike.exe**. Eraldi installeerimisprogrammi ei ole.

Windows7 ja Vistaga töötab ilusti, XP-s jms võib olla transkriptsioonimärkidega probleeme.

# Päringud Liivikesega märgendatud murdekorpuse failidest

Menüüst: **Päringud > Päring mitmest failist**.

Nupu **Lisa** abil saab valida failid, millest otsida. Avanevas dialoogis on võimalik Ctrl või Shift abil valida mitu faili korraga. **Eemalda** kustutab faililoendist liigse faili (selle faili, mis on nupule vajutamise ajal valitud).

Otsitakse lausungeid/lauseid. Mitme sõne otsimiseks samast lausest peab olema midagi valitud lahtrisse **Samas lausungis**. Lahtris **mitte kaugemal kui** tuleb sel juhul määrata, mitu sõna (sh pausid) võib maksimaalselt jääda kahe otsitava sõna vahele. Kui sobivat sõnet on lauses mitu korda, on seda lauset ka väljundis mitu korda. Kuid teise sõne otsingul lausest katkestatakse otsing pärast lähima sobiva sõne leidmist (nii eest- kui ka tagantpoolt), st kui on lause „mina otsin, ootan ja igavlen“ ning pärida „mina“ ja samas lauses peab olema ka „verb + sg.1.“, siis on lauset väljundis üks kord ja leitud sõnedena on märgitud „mina“ ja „otsin“; aga kui niisugusest lausest otsida „verb + sg.1.“ ja „mina“ siis on lauset väljundis kolm korda. Mitme sõna otsing samast lausest on optimeerimata ja üpriski aeglane (mida pikem lubatud distants, seda aeglasem).

Iga välja puhul saab valida, kas otsitakse **täpset vastet** või välja **algusest, keskelt** või **lõpust**. Nt kui kirjutada otsitava sõne lahtrisse „me“ ja valida, et otsitakse lõpust, siis leitakse tekstidest kõik *me-*ga lõppevad sõnad. (Kui lisaks valida, et otsitakse verbe ja mitmuse 1. pöördes, siis leitakse *me-*lõpulised 1. pöörde vormid.) Otsimisel ei eristata suur- ja väiketähti.

Korpuses kasutatavad sõnaliigilühendid on järgmised:

verb V
abiverb Va
eituss. Liitvormis Mn
substantiiv S
pärisnimi H
prosubstantiiv ProS
prepositsioon Pre
postpositsioon Post
adverb Adv
afiksaaladverb AdvA
proadverb ProAdv
modaaladverb ModAdv
adjektiiv A
proadjektiiv ProA
võrdluss.liitsuperlatiivis Ms
järgarvsõna Numj
põhiarvsõna Nump
pronumeraal ProNum
konjunktsioon Konj
küsisõna Intr
diskursusepartikkel Par
interjektsioon Intj
onomatopoeetia Ono
suhtlussõna Suht

Sõnaliigi lahtrisse võib kirjutada ka ainult osa lühendist, nt *num*. Kui siis lahtri kõrval olevast loendist valida *Keskel*, leitakse nii kõik kirjed, kus sõnaliigi väli sisaldab tähejada *num* (st järgarvsõnad, põhiarvsõnad ja pronumeraalid). Kui valida *Alguses,* siis leitakse ainult *numj* ja *nump.*

Otsida saab ka **kommentaaride** järgi. Kuna Liivikese otsingu kasutamiseks peaks vajalikud failid olema oma arvutisse salvestatud, siis võib neile vabalt üldistest kokkulepetest erinevaid märgendeid ja ka kommentaare lisada, just nii, nagu enda uurimistööks vaja.

**Väljundfail** võib olla kas XML või TXT ja Liivikeses seda avada ei saa.

Leitud sõnede ette võib panna suvalise märgi (mis ehk hõlbustab väljundfaili edasist töötlemist unix-is). Kui on kavas väljundfaili töödelda Excelis, võib leitud sõnede ette panna **\t** (st tabulaatori; töötab kui salvestada txt-na).

Iga rea algusesse lisatakse faili ID ja kõneleja ID. Kui valida **lisa rea algusesse täpne faili info**, siis lisatakse ka helifaili info ja murre, kihelkond ning küla. Kui valida **lisa leitud sõne info iga rea algusesse**, siis ka leitud sõne, selle lemma jne.

Lauses saab iga **sõna järele** lisada **märgenduse**. Tekstifailis on see kujul *sõne(M\_lemma\_T\_tähendus\_L\_sõnaliik\_V\_vorm\_K\_kommentaar)*
nt
*pahema(M\_pahem\_T\_vasak\_L\_A\_V\_sg.gen.cmp.\_K\_)*

BOMi lisamine faili algusesse aitab Windowsi programmidel aru saada, et tegemist on UTF-8 kodeeringus failidega. Aga BOM pole kohustuslik.

Kui valid **Lisa XML märgendus**, võiks faili ka salvestada .xml laiendiga. Selliste failide vaatamiseks võiks failiga samas kataloogis olla fail *liivike\_paring.xsl.* Siis vaikimisi avatakse xml-fail internetibrauseris, aga ka (uuem?) Word oskab selliseid faile õigesti kuvada. Wordis on nende failide vaatamiseks kaks varianti, mille vahel faili avades lastakse valida: (a) ainult tekst, (b) liivike\_paring.xsl. Esimene näitab lihtsalt kõiki sõnu üksteise all, teine xsl-stiililehe järgi vormistatud teksti. Kui faili *liivike\_paring.xsl* pole muudetud (muidugi võib seda faili muuta!), paistab tekst Wordis niisugune (päringuks oli täpne sõnekuju *sie*):

sõne: sie [Täpne vaste]

**AVI\_Kaarel\_Haav\_EMH392\_synt@KJ** **Ida/Avi/Kärasi** **sie**, see*, ProS, sg.nom.* ise ise ProS sg.nom. `mõtlesin mõtlema V pers.ind.ipf.sg.1. et et Konj nooh nooh Par (.) et et Konj kui kui Konj (.) sie see ProS sg.nom. onn olema V pers.ind.pr.sg.3. = et et Konj (.) miss mis Intr sg.nom. sie see ProS sg.nom. siis siis ProAdv onn olema V pers.ind.pr.sg.3. (.) muut muu ProS sg.part. kui kui Konj surm surm S sg.nom. (...) jaah jaah Par sie see ProS sg.nom. õhu õhk S sg.gen. vahettus vahetus S sg.nom. (...) nuoh noh Par (.) tul'in tulema V pers.ind.ipf.sg.1. kodu kodu S sg.ill. (.) `riäksin rääkima V pers.ind.ipf.sg.1. naisele naine S sg.all. ära ära Adva se see ProS sg.nom. lugu lugu S sg.nom. (.) et et Konj täma tema ProS sg.nom. kirjuttas kirjutama V pers.ind.ipf.sg.3. rohod rohi S pl.nom. = ja ja Konj (...) naine naine S sg.nom. `akkas hakkama V pers.ind.ipf.sg.3. `nutma nutma V sup. = t et Konj (.) siis siis ProAdv õl'i olema V pers.ind.ipf.sg.3. viel veel Par nuored noor A pl.nom. et et Konj no no Par `kahju kahju S sg.nom. `lahkuda lahkuma V inf. väga väga Adv (.) us'ina usin A sg.gen. (...) jaah jaah Par (...) maa mina ProS sg.nom. = `üt'sin ütlema V pers.ind.ipf.sg.1. = t et Konj eij ei Mn = õle olema V pers.ind.pr.neg. viga viga S sg.part. et et Konj (.) juba juba Par me meie ProS pl.nom. siis siis ProAdv õleme olema V pers.ind.pr.pl.1. koa ka Par viis+kümmend+viis viis+kümmend+viis Nump sg.nom. `aassad aasta S sg.part. vanad vana A pl.nom. (...) mina mina ProS sg.nom. õlen olema V pers.ind.pr.sg.1. (.) siis siis ProAdv me meie ProS pl.nom. õlem olema Va pers.ind.pr.pl.1. = koa ka Par `ulkka hulk S sg.part. `aega aeg S sg.part. juo jo juba Par ühes+kuos ühes+koos Adv elanud elama V nud. = ja ja Konj (...) aga aga Konj = t et Konj (.) nohh noh Par siss siis Par (.) aga aga Konj sie see ProS sg.nom. aeg aeg S sg.nom. läks minema V pers.ind.ipf.sg.3. = edäsi edasi Adva läks edäsi (.) mina mina ProS sg.nom. `pruukkisin pruukima V pers.ind.ipf.sg.1. nied need ProS pl.nom. rohod rohi S pl.nom. = ja ja Konj (...) aga aga Konj sie see ProS sg.nom. rohi rohi S sg.nom. õl'i olema V pers.ind.ipf.sg.3. ia hea A sg.nom. (...) se see ProS sg.nom. rohi rohi S sg.nom. õl'i olema V pers.ind.ipf.sg.3. ia hea A sg.nom. ja ja Konj parandas parandama V pers.ind.ipf.sg.3. minuda mina ProS sg.part. nõnna nõnda ProAdv = et et Konj (...) mull mina ProS sg.ad. = õl'i olema V pers.ind.ipf.sg.3. siis siis ProAdv õl'i olema V pers.ind.ipf.sg.3. (.) nagu nagu Par `ingamises hingamine S sg.in. õl'i olema V pers.ind.ipf.sg.3. (.) `raske raske A sg.nom. (.) `ingamine hingamine S sg.nom. õl'i olema V pers.ind.ipf.sg.3. `raske raske A sg.nom. ma mina ProS sg.nom. kardin kartma V pers.ind.ipf.sg.1. = et et Konj (.) et et Konj tiis'ik tiisikus S sg.nom. `akkab hakkama V pers.ind.pr.sg.3. (...) aga aga Konj (...) ar'st arst S sg.nom. `voattas vaatama V pers.ind.ipf.sg.3. ära ära Adva `ütleb ütlema V pers.ind.pr.sg.3. = et et Konj eij ei Mn = õle olema V pers.ind.pr.neg. (...) eij ei Mn = õle olema V pers.ind.pr.neg. (...) südames süda S sg.in. on olema V pers.ind.pr.sg.3. viga viga S sg.nom. sull sina ProS sg.ad. (...) jaah jaah Par (...) südames süda S sg.in. on olema V pers.ind.pr.sg.3. viga viga S sg.nom. (...) nuo no Par ja ja Konj nuoh noh Par (...) nuo no Par siss siis Par (...) `pruukkisin pruukima V pers.ind.ipf.sg.1. neid need ProS pl.part. `rohtosid rohi S pl.part. `ütleb ütlema V pers.ind.pr.sg.3. (.) nüt nüüd Par (.) pruugi pruukima V pers.imp.pr.sg.2. nied need ProS pl.nom. rohod rohi S pl.nom. ära ära Adva (...) `este esiti Adv (.) siss siis ProAdv pia pidama V pers.imp.pr.sg.2. kaks kaks Nump sg.nom. nädalad nädal S sg.part. vahet vahe S sg.part. (...) pia pidama V pers.imp.pr.sg.2. kaks kaks Nump sg.nom. nädalat nädal S sg.part. vahet vahe S sg.part. (.) siis siis ProAdv (.) too tooma V pers.imp.pr.sg.2. jälle jälle Par uus uus A sg.nom. `ports'jon portsjon S sg.nom. (...) proovi proovima V pers.imp.pr.sg.2. (.) pruugi pruukima V pers.imp.pr.sg.2. siis siis ProAdv (.) nied need ProS pl.nom. (.) rohod rohi S pl.nom. koa ka Par ära ära Adva (...) sie+samma see+sama ProS sg.nom. rohi rohi S sg.nom. (.) a aga Konj kaks kaks Nump sg.nom. nädalad nädal S sg.part. vahet vahe S sg.part. pidäda pidama V inf. (...) no no Par siis siis ProAdv õl'igi olema+ki V pers.ind.ipf.sg.3. seda+viis'i seda+viisi ProAdv = ja ja Konj (...) aga aga Konj mina mina ProS sg.nom. `praavisin praavima paranema V pers.ind.ipf.sg.1. tubliste tublisti Adv (...) `iaste hästi Adv (...) `akkasin hakkama V pers.ind.ipf.sg.1. `tüele tööle Adv `akkasin `tüele (.) aga aga Konj = t et Konj seda see ProS sg.part. `kielas keelama V pers.ind.ipf.sg.3. ära ära Adva et et Konj (.) `rasket raske A sg.part. = tüed töö S sg.part. ära ära Mn pers.imp.pr.sg.2. tie tegema V pers.imp.pr.sg.2.neg. (...) `rasket raske A sg.part. = tüed töö S sg.part. ära ära Mn pers.imp.pr.sg.2. tie tegema V pers.imp.pr.sg.2.neg. (...) nuoh noh Par (.) ja ja Konj sält sealt ProAdv ma mina ProS sg.nom. sain saama V pers.ind.ipf.sg.1. aru aru S sg.part. ma sain aru ku kui Konj ma mina ProS sg.nom. `rasket raske A sg.part. = tüet töö S sg.part. = tegin tegema V pers.ind.ipf.sg.1. (.) **sie** see ProS sg.nom. õl'i olema V pers.ind.ipf.sg.3. mulle mina ProS sg.all. alb halb A sg.nom. siiss siis ProAdv `akkas hakkama V pers.ind.ipf.sg.3. `ingamine hingamine S sg.nom. (.) `raskemast raske A sg.tr.cmp. menema minema V sup.

**AVI\_Kaarel\_Haav\_EMH392\_synt@KJ** **Ida/Avi/Kärasi** **sie**, see*, ProS, sg.nom.* ma mina ProS sg.nom. = i ei Mn tia teadma V pers.ind.pr.neg. kas kas Intr = mul mina ProS sg.ad. on olema V pers.ind.pr.sg.3. `praega praegu Adv **sie** see ProS sg.nom. (.) ar'm arm S sg.nom. piab pidama V pers.ind.pr.sg.3. alles alles Adv õlema olema V sup. (...)

Leitud sõned on punased, rea alguses hallil taustal on eespool kirjeldatud info (kui see on väljundfailis olemas). Kui see fail oleks brauseris avatud, siis näeks sõnedel klõpsates sõnede märgendust. Wordis näeb sõnede märgendust, kui vajutada nupuribal peidetud vormingutähiste näitamise nupule (joonisel punase noolega tähistatud). (Peaks töötama ka siia faili kopeeritud näitega.)



Kui salvestada fail **TXT-**na, siis üks võimalus on seda vaadata Excelis. See variant sobib eriti hästi, kui iga rea algusesse on lisatud leitud sõna info ja leitud sõne ette tabulaator (\t), sest Excelis saab siis faili selle info järgi sortida, sagedusi arvutada jms. Aga Excelis ei tuleks (oleneb Exceli versioonist!) faili mitte lihtsalt avada, vaid importida. Teksti import on Excel 2010 siin (Andmed > Tekstist; varem oli vist Fail-menüüs):



Valides eraldajatega (tabulaator ehk tabeldusmärk) teksti ja kooditabeliks UTF-8, saad sellise faili:



# Liivikese tööaknad

Akende suurusi saab muuta.

### Teksti aken

Teksti aken on kõige ülemine. Selles näidatakse märgendatavat teksti.

Teksti saab kerida. Töötavad ka klahvikombinatsioonid Ctrl+Home ja Ctrl+End.

Kui soovid mingi sõna kohta infot näha või sisestada, klõpsa hiirega sellel sõnal. Sõna, mille kohta alumistes akendes infot näidatakse, on punane.

Sõna kohta sisestatud infot näeb ka siis, kui hiire kursoriga sellel sõnal peatuda.

Lausete vahel on suurem vahe kui tavaridade vahel.

Pealkirjad on paksus kirjas, märkused väiksemas kirjas ja esiletõstud allajoonitud.

Sõnad, mida pole vaja lemmatiseerida (kommentaarid, intervjueerija tekst, vahemärgid jms) on kursiivis. Kui intervjueerija tekst ei ole veel märgendiga „ei märgenda, sest … intervjueerija“, tuleb see lisada. Samuti peab pausimärkidel, muukeelsetel lõikudel jne olema vastav märgend.

Kui teksti sisestamisel on tehtud **vigu**, saab neid muuta, vt menüüst Redaktor -> ...

Avatud teksti kohta sisestatud **infot** saab vaadata/muuta menüüst Tekst -> Teksti info (s.o see info, mis paikneb teksti päises). Tekstiga seotud helifail peaks olema mp3 formaadis.

### Sõneinfo aken

See on keskmine (helesinise taustaga) aken, kus näidatakse **ainult seda infot, mis juba on faili meelde jäetud**. Meeldejätmine pole sama mis faili salvestamine. Analoogselt MS Wordiga: Word jätab tehtud muutused meelde, aga salvestab need alles siis, kui käsid seda teha. Liivikeses faili **salvestamiseks** menüüst Tekst -> Salvesta või CTRL+S.

### Info sisestamise aken

See on alumine aken. Sellesse aknasse sisestatakse kogu vajalik info sõne kohta.

Programmi täpne käitumine sõltub seadistusest (vt edaspidi).

Kui muudad märksõna lahtris olevat infot või valid lemma selle lahtri all olevast loendist, otsib programm ka uuele lemmale vastavad tähendused ja sõnaliigid. Kui muudad tähendust, otsib uue sõnaliigi. Lemma soovitamisel arvestab sõnet, tähenduse soovitamisel arvestab lemmat, sõnaliigi soovitamisel arvestab lemmat ja tähendust.

Liivi ja läti tähti, mida eesti tähestikus/klaviatuuril pole, saab sisestada kahte moodi: (1) kas vastavate nuppudega või (2) tähekombinatsioonide abil. Kui soovid eelmisele tähele lisada *cedilla*, alla või üles koma, tipi tähe järele koma; kui soovid lisada tähe peale kriipsu; tipi tähe järele kriips, samamoodi punkt tähe kohale, punktid (koolon) tähe kohale, katus tähe kohale. Kui soovid sisestada nt sidekriipsuga kokku kirjutatud liitsõna, võta ära linnuke „Asenda märgid tippimise ajal“ eest. Neid märke saab lisada ainult lemma ja läti lahtrisse.

Vormiinfo sisestamiseks avaneb eraldi aken. Ühele tekstisõnale on võimalik korraga valida ka mitu nt käändemärgendit, nt *all* ja *ad* vms, kui teksti põhjal pole võimalik otsustada, kummaga tegu. St programm ei kontrolli, kas märgendid lähevad üksteisega vastuollu. Programm siiski kontrollib, kas kogu vajalik info on sisestatud. Seda näitavad värvid vormiinfo dialoogis: punane – alajaotises on märgend arvatavasti puudu, oranž – alajaotises on märgend arvatavasti liigne, roheline – kõik peaks olema OK.

Kommentaariks võib kirjutada ükskõik mida. Sellesse lahtrisse saab kirjutada tekkinud küsimusi jms, neid saab pärast otsida (et teistega arutada). Murdekorpuse vanades failides märgendiga <fra> tähistatud info läheb ka sellesse lahtrisse teksti importimisel.

Tähendus ja ka kõik muu tuleb lisada ainult konkreetse tekstisõna kohta, st mitte *kumb ‘kes, mis’*, vaid ainult konkreetsesse konteksti sobiv tähendus.

Kirjavahemärgile ega intervjueerija tekstile ei saa märksõna ega midagi lisada (kommentaari saab). St kui mõni sõna on ekslikult kirjavahemärgiks pandud, tuleb kõigepealt see märgend ära muuta ja meelde jätta.

Nupp “Tagasi” liigub eelmisele sõnale, mille kohta infot sisestasid.

Nupp “Jäta meelde” jätab antud sõne kohta sisestatud info meelde. Meeldejätmine pole sama mis faili salvestamine. Analoogselt MS Wordiga: Word jätab tehtud muutused meelde, aga salvestab need alles siis, kui käsid seda teha. Liivikeses menüüst Tekst -> Salvesta või CTRL + S.

Nupp “Näita järgmist” näitab järgmist tekstisõna. Nupu kõrval oleva noolekese abil saab valida, kas liigutakse järgmisele sõnale, järgmisele märgendamist vajavale sõnale, järgmisele samasugusele sõnale vm. Tehtud valik jäetakse meelde ja muutub ka kiri nupul.

Nupp “Jäta meelde ja näita järgmist” jätab antud sõne kohta sisestatud info meelde ja seejärel näitab järgmist sõnet. Millist järgmist näidatakse, sõltub “Näita järgmist” nupu alt tehtud valikust.

Nupp „Lause kontroll“ avab eraldi aknas lause koos sisestatud märgenditega. Nii on kõike sisestatut ehk lihtsam üle vaadata.

Nupp „Lisa lause tõlge“ avab akna, kuhu saab sisestada lause tõlked eesti, läti ja inglise keelde.

Nupp „Märksõna info“ avab märkmefaili vastava märksõna kohalt. Kasutatava märkmefaili saab määrata seadistusdialoogis. Mõeldus selleks, et panna enda jaoks kirja märgendamisjuhiseid, nt et vormi *enese* märksõna on *enese,* mitte *ise* vms.

Akna allservas on näha lause alguskoht (minutid/sekundid) helifailis. Selle muutmiseks või määramiseks vali menüüst „Redaktor“ -> „Lause algus helifailis“ (kiirklahv Ctrl+F) või seadistusdialoogis „Lisan lausetele alguspositsiooni helifailis“ (vt allpool).

**Märkused**

Kui sõnaliik, vorm või keel on -, siis seda ei salvestata, aga kui lemma või tähendus on -, siis salvestatakse, st kuigi lemmade ja tähenduste loendis näed kriipsu, ära seda ise neisse lahtritesse kirjuta (kui kriips on valitud, on ju lahter tühi).

Kui pole tähendust, sõnaliiki, sõnavormi, siis see variant pannakse alati loendi lõppu, aga loendist automaatselt valitakse esimene asi – sel juhul loendist valitud asi pole sama mis failis salvestatu!

# Abisõnastikud

Sõnastikku kasutatakse soovituste otsimisel. Kasutatava sõnastiku saab valida seadistusdialoogis (Töövahendid -> Soovitaja seaded). Sõnastik on kindla struktuuriga XML-fail.

Liivi keele jaoks on Viitso sõnaraamatu põhjal tehtud abisõnastik *liivi\_korpuse\_sonastik.xml.*

Murdekorpuses võiks teha igale murderühmale oma sõnastiku, mingi arusaadava failinimega (nt *murderühm\_abisnstk.xml*) ja hoida seda nende failidega samas kataloogis. Märgendama asudes on oluline kontrollida, et kasutataks õiget sõnastikku.

### Sõnade lisamine sõnastikku

Sõnu saab lisada nii uude (tühja) kui ka mõnda juba olemasolevasse sõnastikku.

Sõnastikku saab lisada ainult avatud teksti sõnu. Selleks tuleb valida menüüst Tekst -> Lisa avatud teksti sõnad sõnastikku. Sõnastikku ei lisata sõnesid, millel pole lemmat. Seda funktsiooni on mõtet kasutada siis, kui on olemas varem märgendatud tekste, mille põhjal abisõnastik teha.

Jooksvalt, teksti märgendamise ajal sõnade lisamiseks sõnastikku tuleb seadistusdialoogis vastav valik teha.

### Sõnastiku redigeerimine

Kui kogemata on sõnastikku sattunud vigu, mida programm siis edaspidi muudkui soovitab, saab sõnastikku parandada.

Töövahendid -> Redigeeri sõnastikku. Sõnastiku alguses esimene sõne, lemma ja ühend on --, seda ei tohi ära kustutada (ega ei saagi praegu).

Sõnastiku redigeerimisel saab vormiinfot loendist valida, aga saab ka ise midagi muud kirjutada.

Sõnastiku redigeerimisdialoogis kustutamisnupp (X). Sõnastiku muutmisel peab arvestama, et kui nt tähendust muudad ja vajutad ‘Lisa’, jääb vana variant ka alles. See vana variant tuleb eraldi ära kustutada.

# Seadistus

Seadistusdialoogi vaatamiseks vali menüüst Töövahendid -> Soovitaja seaded.

**Liivikese soovitaja** otsib, kas samakujulisele sõnele on avatud failis juba infot lisatud või kas selline sõne on olemas sõnastikus. Leitud infot näidatakse sisestamisakna loendites. Sõnede võrdlemisel on võimalik ignoreerida sõnalõppe (nt kui tekstis varem sõnele *maja* on pandud lemmaks *maja* ja ignoreeritakse lõppu *sse*, pakutakse lemmat *maja* ka sõnale *majasse*).

Ignoreeritavate sõnalõppude loendit saab seadistusdialoogis muuta. See loend jäetakse meelde, nii et programmi järgmisel käivitamisel on see sama.

**Liivi keele abi** all on valik „Soovita morf vormi sõnalõpu järgi“. Nende soovitused on jällegi Viitso sõnaraamatu järgi tehtud.

"Livonian Language Tools" – analüüsib tekstisõna lehel <http://giellatekno.uit.no/cgi/d-liv.eng.html> oleva tarkuse abil (vajalik internetiühendus). Kui lehe autor oma programmi väljundis midagi muutab, võib tekkida ootamatuid tõrkeid ;)

**Homonüümide kontrolli** automaatsetvõimalust pole.

**Läti** ja **inglise** vastete soovitamise saab ära keelata.

Kui XML faili päises on kirjas helifaili nimi, siis võiks ka iga lause siduda positsiooniga selles helifailis. Et seda märgendamise ajal automaatselt küsitaks (mitmendal minutil/sekundil helifailis lause algab), vali **Lisan lausetele alguspositsiooni helifailis.**

Seadistusdialoogis tehtud valikud jäetakse meelde ja kui järgmisel korral programmi käivitad, kasutatakse vaikimisi samu seadeid.

# Otsing

Menüüst **Otsing -> Otsi**. Kui teed märksõna, tähenduse, sõnaliigi või millegi muu ette ainult linnukese, aga lahtri jätad tühjaks (või paned rippmenüüga lahtrisse -), otsitakse sõnesid, millel pole märksõna, tähendust, sõnaliiki vm. Kui kirjutad lahtrisse \* (või paned rippmenüüga lahtrisse *- suvaline -*), otsitaksegi suvalist märksõna, tähendust, sõnaliiki vm.

Otsida võib ka nt sõnet *pea*, mille märksõnaks on *pidama*, st asju võib kombineerida.

Kui valid variandi, kus leitud sõnad tähistatakse värviga, saad järjest teha mitu otsingut (erineva värviga) ja pärast ainult mööda värvilisi sõnu liikuda. Otsing -> Tühista senised otsingud korjab sel juhul sõnadelt värvilised taustad ära. Nii saab nt otsida eestikeelseid sõnu, millele veel pole märksõna lisatud, ja ainult mööda neid liikuda. (Sisestamisaknas nupu “Näita järgmist” kõrval olev nooleke abiks või Otsingu-menüü kaks viimast käsku.)

Otsing -> Jäta meelde praegune asukoht jätab meelde, millisel sõnal asud. Nii saad nt info sisestamise vahepeal otsida, kas samasugust sõnet edaspidigi esineb, ja hiljem esialgse sõne juurde tagasi pöörduda (Otsing -> Mine meelde jäetud asukohta). Meelde jäetud asukoht unustatakse, kui sulged programmi või avad uue faili.

**Ebatäpne** **otsing** on ka olemas.

Saab ka **otsida ja asendada**, nt kui oled otsustanud mõnd märgendust igal pool muuta.

Otsida saab ainult avatud teksti piires.

# Nupud

****

# Statistika

(Menüüst) näitab, kui palju on tekstis sõnu ja kui palju neist on märgendatud.

Et sama asja visuaalselt näha, ilma konkreetsete arvudega, võib teha otsingu, mis värvib kõik märksõnata sõned tekstiaknas nt kollaseks.

# Sisendtekst

### Formaat

Sisendtekst on *unicode*’is tekst (nt MS Wordis salvestamisel valida *Save as Type: Unicode Text (\*.txt)*.

### Märgendus

Märgendus oleneb tekstitüübist, kas suuline või kirjalik.

**Transkribeeritud** (st siis suulise) teksti alguses võiks olla märgend <info>. Kui seda pole, algab programmi jaoks kohe sisu. Märgendite <com>...</com> vahel olev läheb sisu alla.

Kui teksti alguses (mingeid tühikuid ega tühje ridu ei tohi eelneda!) on <info>...</info> (nende märgendite vahel peaks olema kirjas keelejuhid, litereerijad, aastaarvud jm oluline taustainfo), võivad selle järel olla ka

<kla>...</kla> või <kyla>...</kyla> *küla*

<plk>...</plk> või <pealkiri>...</pealkiri> *pealkiri*

<a>...<a> või <aasta>...</aasta> *aasta*

<kj>...<kj> või <keelejuht>...</keelejuht> *keelejuht/keelejuhid*

Kõiki muid välju ignoreeritakse. Lihtsam on teksti importimisel vastav info ühest lahtrist teise kopeerida kui TXT faili käsitsi märgendada.

Fail peab algama märgendiga!

Sisuteksti alguses peab olema kas märgend <com> või <who=?> või <u who=?>

Pole vaja vooru lõppu </u> märgendit, kui iga voor algab uuelt realt.

**Väga pika vooru võiks jagada mitmeks (reavahetustega).** Seda automaatselt ei tehta. „Lauseid“ saab tükeldada ka märgendamise ajal (Menüüst **Redaktor > Tükelda** lause, või sama käsk hiire parema nupuga).

Märgendi ja sõna vahel võib olla tühik, aga ei pea.

Programm tunneb ära järgmised kirjavahemärgid (neid pole vaja sõnadest tühikutega eraldada, kõik muud on vaja mõlemalt poolt tühikutega eraldada):

. , : : ! ? ( ) [ ] " =

— “ ” „ ‟ ‿

Suulistes tekstides, olenevalt transkriptsioonitüübist, = ja ‿ märgivad kokkuhääldust. Erinevad sõnad loetakse XML-faili eraldi sõnadena. + märgib liitsõnu või liiteid. Liitsõnu XML failis ei lahutata.

<com>kommentaar</com>

Keelejuht ei saa vahetuda vooru keskel.

**Kirjalikes tekstides:**

Iga lausekirjutada eraldi reale.

Leheküljevahetustmärgib **<lk=X>** , kus X-i asemel on algava lk number. Märgend peab olema ka faili alguses.

Nt. euch aber die jhr Trübsal leidet / <lk=22> ruhe mit vns / wann nun der HErr Jesus wird offenbahret werden von Himmel .

<lk=21> Sedda <lk=22> peame meije tehhalle pannema / ninck mitte nende pilckjade kahs meije omma hendesse himmo perrast keima / ninck ütlema / kussa on se towotus temma tullemissest ?

**<p>** pealkiri **</p>**

**<m>** märkus **</m>**

**<a>** esiletõst **</a>**

Koos ei saa esineda

1. pealkiri, märkus, esiletõst
2. kommentaar ja muukeelne (muukeelne on mõeldud, kui vestluses osalejad vahetavad keelt; kommentaar on naermine jms situatsiooni kohta käiv lisainfo

Sellise TXT faili avamiseks programmis tuleb valida Tekst -> Impordi tekst. Ilmuvas faili info dialoogis sisestada nõutud info. Loenditest saab valida, aga loenditega lahtritesse saab ka ise mingit muud teksti sisestada, ei pea kindlasti olemasolevate variantide seast valima.

### Sisestusvead

Sisestamisel tehtud vigu saab parandada, vt menüüst **Redaktor**.

# Väljundtekst

### Formaat

XML (versioon 1.0, kodeering UTF-8)

### Märgendus

Üldine põhimõte, et atribuutidena esitatakse see info, mida saab antud elemendil olla ainult üks. Päises siiski pole päris nii.

<?xml version="1.0" encoding="UTF-8"?>

<?xml-stylesheet type="text/xsl" href="liivike\_tekst.xsl"?>

<liivike>

 <info>

 <tyyp>suuline</tyyp>

 <teema>TEEMA</teema>

 <keel>KEEL</keel>

 <murre>MURRE</murre>

 <kirjaviis>IPA</kirjaviis>

 <muu>MUU INFO</muu>

 <lindistus lindistaja="LINDISTAJA NIMI" aasta="AASTA" fail="misiganes.mp3">ARHIIVIVIIDE</lindistus>

 <litereering litereerija="LITEREERIJA NIMI" aasta="AASTA">ARHIIVIVIIDE</litereering>

 <kyla longituud="1.2" latituud="3.4">KÜLA</kyla>

 <teos pealkiri="PEALKIRI" aasta="AASTA" koht="KOHT">AUTOR</teos>

 <keelejuht id="KJ1" haridus="MAG" synniaasta="1800" synnikoht="RAKVERE" vanus="100">KEELE JUHT</keelejuht>

 <keelejuht id="KJ2" haridus="" synniaasta="" synnikoht="" vanus="">TORE INIMENE</keelejuht>

 <intervjueerija id="OU" haridus="" synniaasta="" synnikoht="" vanus="">ONU UNO</intervjueerija>

 </info>

 <sisu id="ID">

 <lause id="l1" koneleja="KJ1" helipos="0">

 <sone id="l1\_s1" lemma="see" tahendus="too" lati="läti" ingl="that" liik="ProS" vorm="sg.nom."

 stiil="märkus" kommentaar="kommentaar">selless</sone>

 <sone id="l1\_s2" meta="vahemärk">(.)</sone>

 <lause id="l2"><sone id="l2\_s1" meta="kommentaar">katkestus</sone></lause>

 <lause id="l3" koneleja="KJ2" helipos="6"><sone id="l3\_s1">.</sone></lause>

 </sisu>

</liivike>

# Murdekorpuse failide importimine

Teeb algsest XML-failist

1. vahefaili (failinimes juures [1], kus on parandatud märgendus ja mis on ka UTF-8)
2. liivikesele sobiva UTF-8 XML faili (failinimes juures [UTF-8])

St, algset faili ei muuda.

Tekstiinfo tabelisse peaks kindlasti kohe kirjutama õiged intervjueerijate ID-d (st need lühendid/initsiaalid, mida on failis kasutatud). Ainult sel juhul märgendatakse intervjueerija tekst automaatselt intervjueerija tekstiks ja saab kontrollida, kas keelejuhi tekstis on kõik sõnad lemmatiseeritud.

Enne Teksti info dialoogi ilmumist võib suuremate failide puhul kuluda ärritavalt palju aega!

Importimise ajal leitud kahtlased kohad on värviliselt märgitud (kommentaari lahtris „import: viga“). Kahtlaseks peab programm märgendeid, mida uues versioonis ei tohiks kasutada, aga ka seda, kui osa vajalikust infost on puudu. Kui vajadusel parandus tehtud, tuleks kommentaar „import: viga“ ka ära kustutada.

Vanas ja uues versioonis on mõned märgendid erinevad (ja morf info märgendites tühikute asemel punktid):
n -> nom
g -> gen
p -> part
ps -> pers
v -> vpts
Mn on ainult sõnaliigimärgend, mitte morf.vorm.

Sõnaliikide märgendite hulgas pole enam Ak ja As (nende puhul on sõnaliik A ja vormiinfos cmp või sup)

**Kuidas võiks neid tekste importida?**

(Kõigepealt kopeeriksin kõik imporditavad failid uude kataloogi)

1. Valida/luua seadistusdialoogis õige sõnastik, kui seda veel pole
2. Tekst -> Impordi vana murdekorpuse tekst
3. Sisestada korrektne info teksti kohta
4. Vaadata üle programmi leitud kahtlased kohad (Abiks nupp *Näita järgmist värviga tähistatud sõna* (vajuta noolekesele nupu *Näite järgmist* kõrval)
5. Salvesta fail
6. Tekst -> Lisa avatud teksti sõnad sõnastikku
7. Impordi järgmine sama murde tekst

(Lõpuks, kui kõik näis toimivat, võib algfailid ja vahefailid ära kustutada.)

# XSL-id

## Liivike\_paring

<?xml version="1.0"?>

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

<xsl:template match="/">

<html>

<head>

<style type="text/css"><![CDATA[

<!--

BODY{MARGIN: 30px; FONT-WEIGHT: normal; FONT-SIZE: 14px; COLOR: black; FONT-FAMILY: Arial; BACKGROUND-COLOR: white}

.lemma{FONT-WEIGHT: bold }

.liik{FONT-SIZE: 12px; FONT-WEIGHT: bold; TEXT-TRANSFORM: uppercase }

.vorm{FONT-SIZE: 12px }

.sone{color: black; background-color: white; CURSOR: pointer }

.leitudsone{color: red; background-color: white; CURSOR: pointer }

.kommentaar{COLOR: #b22222 }

.lause{font-size: 12px; background-color: #dbdbdb }

.heli{font-size: 12px; background-color: #f0f0f0 }

.koht{font-size: 12px; background-color: #f0f0f0 }

.sona{font-size: 12px; background-color: #f0f0f0 }

.pais{font-size: 13px; COLOR: #003ca0}

.soneinfo{ color: black; background-color: #ffffc4; display: none }

.juhend{font-size: 10px; COLOR: #000000 }

A{ font-size: 11px; COLOR: #002cb9; TEXT-DECORATION: none }

A:hover{ COLOR: #e86004 }

-->

]]></style>

<script type="text/javascript" language="javascript"><![CDATA[

 function ShowHide(Parent){

 var Object=Parent.childNodes[Parent.childNodes.length-1];

 if (!Object) return;

 if (Object.style.display!='inline') Object.style.display='inline';

 else Object.style.display='none';

 }

 ]]></script>

</head>

<title>Päring murdekorpusest</title>

<body>

<span class="juhend">Märgenduse vaatamiseks klõpsa soovitud sõnel </span><br/>

<span class="pais"><xsl:value-of select="liivike-paring/@otsitav"/></span><br/>

<xsl:for-each select="liivike-paring/lause">

 <p>

 <span class="lause">

 <b><xsl:value-of select="./@tekst"/><xsl:text>@</xsl:text>

 <xsl:value-of select="./@koneleja"/></b>

 <xsl:text> </xsl:text>

 </span>

 <xsl:if test="./@helifail!=''">

 <span class="heli">

 <b><xsl:value-of select="./@helifail"/><xsl:text>@</xsl:text>

 <xsl:value-of select="./@helipos"/></b>

 <xsl:text> </xsl:text>

 </span>

 </xsl:if>

 <span class="koht">

 <b><xsl:value-of select="./@murre"/><xsl:text>/</xsl:text>

 <xsl:value-of select="./@kihelkond"/><xsl:text>/</xsl:text>

 <xsl:value-of select="./@kyla"/></b>

 <xsl:text> </xsl:text>

 </span>

 <span class="sona">

 <b><xsl:value-of select="./@sone"/></b>

 <xsl:if test="./@lemma!=''"><xsl:text>, </xsl:text><xsl:value-of select="./@lemma"/></xsl:if>

 <xsl:if test="./@tahendus!=''"><xsl:text>, '</xsl:text><xsl:value-of select="./@tahendus"/><xsl:text>'</xsl:text></xsl:if>

 <xsl:if test="./@liik!=''"><i><xsl:text>, </xsl:text><xsl:value-of select="./@liik"/></i></xsl:if>

 <xsl:if test="./@vorm!=''"><i><xsl:text>, </xsl:text><xsl:value-of select="./@vorm"/></i></xsl:if>

 <xsl:if test="./@kommentaar!=''"><xsl:text> (</xsl:text><xsl:value-of select="./@kommentaar"/><xsl:text>)</xsl:text></xsl:if>

 <xsl:text> </xsl:text>

 </span>

 <xsl:for-each select="sone">

 <xsl:choose>

 <xsl:when test="@tyyp='otsitav'">

 <span class="leitudsone" onclick="ShowHide(this)">

 <b><xsl:value-of select="."/><xsl:text> </xsl:text></b>

 <span class="soneinfo">

 <span class="lemma"><xsl:value-of select="@lemma"/><xsl:text> </xsl:text></span>

 <span class="tahendus"><xsl:value-of select="@tahendus"/><xsl:text> </xsl:text></span>

 <span class="liik"><xsl:value-of select="@liik"/><xsl:text> </xsl:text></span>

 <span class="vorm"><xsl:value-of select="@vorm"/><xsl:text> </xsl:text></span>

 <span class="kommentaar"><xsl:value-of select="@kommentaar"/><xsl:text> </xsl:text></span>

 </span>

 </span>

 </xsl:when>

 <xsl:otherwise>

 <span class="sone" onclick="ShowHide(this)">

 <xsl:value-of select="."/><xsl:text> </xsl:text>

 <span class="soneinfo">

 <span class="lemma"><xsl:value-of select="@lemma"/><xsl:text> </xsl:text></span>

 <span class="tahendus"><xsl:value-of select="@tahendus"/><xsl:text> </xsl:text></span>

 <span class="liik"><xsl:value-of select="@liik"/><xsl:text> </xsl:text></span>

 <span class="vorm"><xsl:value-of select="@vorm"/><xsl:text> </xsl:text></span>

 <span class="kommentaar"><xsl:value-of select="@kommentaar"/><xsl:text> </xsl:text></span>

 </span>

 </span>

 </xsl:otherwise>

 </xsl:choose>

 </xsl:for-each>

 </p>

</xsl:for-each>

</body>

</html>

</xsl:template>

</xsl:stylesheet>