

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Kaarel Sõrmus
Kantseliidi- ja paronüümituvastaja
Bakalaureusetöö (9 EAP)

Juhendaja: Krista Liin

Tartu 2017

Kantseliidi- ja paronüümituvastaja

Lühikokkuvõte:

Hea suhtlemise aluseks on üheselt mõistetav keel. Kantseliitliku keelekasutuse vältimine ja paronüümide ehk kergesti segi aetavate sõnade õigesti kasutamine muudab teksti selgemaks ja arusaadavamaks. Bakalaureusetöö eesmärk oli luua programm, mis suudab eestikeelsest tekstist üles leida paronüümid ja kantseliidi ning need tekstis esile tuua, kuna teadaolevalt eesti keele jaoks selline programm puudub. Töös kirjeldatakse kantseliidi ja paronüümi olemust ning tuuakse välja, miks on tähtis kantseliiti vältida ja paronüüme õigesti kasutada. Samuti antakse ülevaade sarnastest välismaistest programmidest. Töös kirjeldatakse, kuidas programm paronüüme leiab ning kuidas ja milliseid kantseliitlikke konstruktsioone otsib.

Võtmesõnad: paronüüm, kantseliit, morfoloogiline analüüs, süntaktiline analüüs

CERCS: P175 Informaatika, süsteemiteooria

Officialese and paronym detector

Abstract:

Good communication is based on unambiguous language. Avoiding officialese and using paronyms correctly makes text clearer and more understandable. The aim of this Bachelor's Thesis is to make a program which can find officialese and paronyms from Estonian text and highlight them, because no known program like this exists for Estonian language. Thesis describes what are paronyms and officialese and explains why it is important to use paronyms correctly and to avoid officialese. Thesis also gives examples of similar programs for foreign languages. The final section describes how the created program detects paronyms and officialese and which forms of officialese it is able to detect.

Keywords: paronym, officialese, morphological analysis, syntactic analysis

CERCS: P175 Informatics, systems theory

Sisukord

Sissejuhatus.....	5
1. Kantseliit.....	6
1.1 Kust tuleb kantseliit?.....	6
1.2 Kantseliidi õigesti kasutamise tähtsus.....	7
2. Paronüümid.....	8
2.1 Paronüümide õigesti kasutamise tähtsus.....	8
3. Selge keele liikumine.....	9
4. Sarnased programmid.....	10
4.1 Microsoft Word.....	10
4.2 OpenOffice – LanguageTool.....	11
5. Programmi ülevaade.....	13
5.1 EstNLTK.....	13
5.2 Tkinter.....	13
5.3 Programmi kasutajaliides.....	14
5.4 Paronüümide märkimine.....	14
5.5 Kantseliidi märkimine.....	15
5.6 Kantseliidi märkimine kasutades lemmatiseerimist ja süntaktilist analüüsi.....	16
5.6.1 Tarbetud sõnad.....	16
5.6.2 Nominalisatsiooni <i>mine</i> -vormid.....	17
5.6.3 Olema + kesksõna.....	19

5.7 Kantseliidi märkimine kasutades morfoloogilist ja süntaktilist analüüsi.....	20
5.7.1 <i>Poolt</i> -tarind.....	20
5.7.2 Määrus saavas käändes.....	21
5.7.3 <i>lt</i> -määrsõnad	22
5.7.4 Mitmus ainsuse asemel.....	23
5.8 Infoaken.....	23
6. Edasiarendamise võimalused	25
Kokkuvõte.....	26
Viidatud kirjandus.....	27
Lisad.....	28
Lisa 1. Programmi käivitamisjuhend	28

Sissejuhatus

Keeletoimetajad jt keelega tegelevad inimesed heidavad ette, et keelde on sünenud liiga palju kantseliiti, mis teeb tekstidest arusaamise hägusaks. Kantseliit on keelekasutus, mida iseloomustab keerukas ning mitmetimõistetav stiil [1]. Ka paronüümide valesti kasutamise tõttu muutub tekst raskemini mõistetavaks. Paronüümid ehk sarnassõnad on kõlalt lähedased, kuid erineva tähendusega sõnad [2].

Kantseliiti ja paronüümide valesti kasutamist esineb kõigis arenenumates keeltes, ka eesti keeles. Kantseliiti ning valesti kasutatud paronüüme esineb nii kirjakeeles kui ka suhtluskeeles [3][4]. Eriti oluline on kantseliiti vältida ja paronüüme õigesti kasutada ametlikes tekstides, kuna need peavad informatsiooni üheselt ja arusaadavalt edasi andma [1].

Paronüümide korrektse kasutuse kohta saab teavet hankida sõnastikest, kus on sarnaselt kõlavate sõnade õiged kasutusviisid välja toodud.[2] Kantseliidi kohta kirjutatud raamatutes seletatakse, milliseid lausekonstruktsioone ja sõnu peaks tekstis vältima. Sageli jätab kirjutaja aga teksti kontrollimata, sest töö sõnaraamatutega on tüütu ja aeganõudev. Sõnastike sirvimise asemel on mugavam kasutada automatiseeritud rakendust. Töö eesmärk ongi luua programm, mis suudab kasutaja sisestatud eestikeelsest tekstist üles leida kantseliiti ning paronüüme, et neile tähelepanu juhtida.

Töö koosneb kahest osast – praktilisest ning teoreetilisest. Teoreetiline osa annab laiemat ülevaadet paronüümidest ning kantseliidist eesti keeles – kust nad tulevad ning miks neid vältima peaks. Kirjutatakse ka selge keele liikumisest ning antakse ülevaade sarnastest programmidest. Praktilises osas kirjeldatakse programmi funktsionaalsust. Praktiline töö on valminud kasutades raamatuid „Väike paronüümisõnastik“ [2] ning „Kuidas hoiduda kantseliidist“ [1].

1. Kantseliit

Kantseliit on keelekasutus, mida iseloomustab keerukas, ebamäärane või mitmeti ja raskesti mõistetav stiil. Laiemalt peetakse kantseliitlikuks tekstiks paljusõnalist ja abstraktset teksti. Kantseliiti on kirjeldatud ka kui bürokraatiaslängi, mis ei kuulu ametlikku keelekasutusse [1]. Eestis tegeleb avaliku keelekasutuse parandamisega Eesti Keele Instituut ning ka 2007. aastal nelja asutuse koostöös loodud Keelehooldekeskus [5]. Keelehooldekeskuse väljaannete hulka kuulub näiteks raamat „Kuidas hoiduda kantseliidist“ [1], mille eesmärk on juhtida tähelepanu keele õigele kasutamisele.

1.1 Kust tuleb kantseliit?

Uno Liivaku on kirjutanud juba 1972. aastal [6]: „Praegu on kogu ametnikkond ühtse riikliku süsteemi teenistuses. Kirjade ja teiste dokumentide sõnastus minetab järjest isikupära. Asutustevahelised sidemed on tihedad. Ühes kantseliis keele kohta langetatud väärotsus jõuab dokumentide vahendusel peagi teistesse, kus keele asjus kriitikavõimetud ametnikud selle omaks võtavad ja seda omakorda levitama hakkavad.“ Kahjuks ei ole tänapäeval midagi muutunud. Bürokraatlikke kirju saadetakse iga päevaga aina rohkem. Kui igapäevaselt saadetakse kantseliiti täis teksti, siis ei hakka ka keeleteadlikumad inimesed teksti õigsuses kahtlema ja nii sulandub kantseliit meie igapäevasesse keelekasutusse [4].

Järgnev on refereeritud Krista Kerge artikli põhja [7], kus osutatakse, et teksti süntaksi keerukus ja abstraktsus on üks-ühele seoses teksti ametlikkusega. Tihti määrab teksti autor oma sõnavara kasutamisega, kas tekst on loetav ainult eriala tundvale inimesele või saab tekstist informatsiooni kätte ka mitteasjatundja. Heaks näiteks on statistikakeel, kus loetakse kantseliidiks teksti keerukust, ümberütlemisi, ületäpsustamisi, stampkeelendeid, võõrõigulisust ning piiratud sõnavara. Statistikakeeles, nagu ka õiguskeeles, kasutatakse liiga palju nominaalstiili – püütakse end väljendada väheste tegusõnade ning paljude nimisõnadega. Sealhulgas kasutatakse selliseid tegusõnu, mis on oma algse tähenduse kaotanud ning nüüd laiuvad ametlikes keeltes ning tekstides, näiteks sõnad: *läbi viima*, *tegelema*, *teostama*, *sooritama* ja *toimuma*. Artiklist selgub, et kuna neid sõnu kasutatakse pea igal pool, siis jääbki tavalisele keelekasutajale väär mulje, et nende tegusõnade kasutamine on hea tava [7].

Kantseliit on tunginud ka suulisesse keelde – tihti kasutavad inimesed avalikel ülesastumistel kantseliitlikku keelt, „ametlikke“ väljendeid [4]. Kasutatakse tühisõnu (sõnad, mis ei lisa lausele mõtet või tähendust), üle-tuletusi (tuletusliidetega liialdamine, liiteid on sõnas rohkem kui mõistmiseks tarvilik) ja liaseid väljendeid (väljend, milles on sisukordus) – seda selleks, et kõne jäta keerulisema ning targema mulje [4]. Kantseliit võib tekkida ka võõra keele lausemallide ning grammatika ülevõtmise kaudu, mistõttu võib eestikeelne lause muutuda raskesti loetavaks ning tihti ka liiga pikaks [8].

1.2 Kantseliidi õigesti kasutamise tähtsus

Eelkõige on vaja kantseliiti vältida tarbekeelsetes tekstides, kuna tarbetekstide peamine eesmärk on infot arusaadavalt edastada. Selleks tuleb vältida kõrvalekaldeid harilikust keelekasutusest ning arvestada keelenorme. Tarbetekstides olev kantseliit võib teksti sisu muuta raskemini mõistetavaks [1].

Siinkohal on paslik tuua näide sama sisuga lausetest, kus esimeses lauses on kasutatud kantseliitlikku keelekasutust [9]:

- 1) Menetlusteenistus on antud piirkonnas juba teostanud pistelist kontrollimist ja kindlasti teostab edaspidi kontrolli märgitud piirkonnas.
- 2) Menetlusteenistus on nimetatud piirkonda juba pisteliselt kontrollinud ja teeb seda kindlasti ka edaspidi.

Keeletundlikuma inimese jaoks kõlab teine lause kindlasti paremini ning selgemalt [9].

Paar näidet ka kantseliiti sisaldavate lausete konstruktsioonidest. Järgmised näited on saadud Reili Arguse tekstist „Kantseliit, mis lööb oimetuks“ [8]:

- 1) Registreerimine toimub raviarstide, patoloogide või kohtuarstide poolt saadetud andmekaartide põhjal sekretär-assistendi poolt vahetult enne üritust.
- 2) Sekretär-assistent registreerib osalejad raviarstide, patoloogide ja kohtuarstide saadetud andmekaartide põhjal vahetult enne üritust.

Esimeses lauses on kasutatud võõrapärast lausemalli – *poolt*-malli. Lause jätab tegija tagaplaanile ning jääb mulje, et registreerimine toimuks ise. Teine lause on korrektne – tegija viiakse lause algusesse ning lause tundub palju loogilisem [8]. Või näiteks:

- 1) Hinnanumbrid selles veini tootesegmendis on ülesmäge minevad.
- 2) Selle segmendi veinide hind tõuseb.

Esimeses lauses kasutatakse ülearust sõna „number“ ning võõrapärast stiili – „miski on missugune“. Teine lause on tähenduselt sama, kuid on palju konkreetsem ning arusaadavam.[8]

2. Paronüümid

Paronüümid ehk sarnassõnad on sõnad, mis on kõlalt ja kirjapildilt lähedased, kuid tähendavad erinevaid asju. Tavaliselt on paronüümid tekkinud samast sõnatüvest [2]. Leidub paronüüme, mis on ka tähenduselt lähedased, ning sõnu, mis erinevad nii minimaalselt, et kõnekeeles neid ei eristata [10]. Keelehooldekeskusel on ka paronüümide kohta väljaanne „Väike paronüümisõnastik“ [2].

2.1 Paronüümide õigesti kasutamise tähtsus

Vales tähenduses sõnade kasutamine võib muuta lause sisu. Suuremat tähelepanu tuleks korrektsele sõnakasutusele pöörata tekstides, mis nõuavad täpsust [2]. Näiteks koolikirjandites kipuvad paronüümidega seotud vead esinema sellepärast, et kõnekeeles ei tehta neil sõnadel tihti vahet – näiteks võib tuua sõnad „kestus“ ja „kestvus“ [11].

Teiseks võib tuua näite samasisulistest lausetest, kus esimeses lauses on autori mõte paronüümide „ohjama“ ja „ohjeldama“ segiajamise tõttu läinud:

- 1) Politsei ohjab kuritegevust.
- 2) Politsei ohjeldab kuritegevust.

Esimene lause tähendab, et politsei juhib kuritegevust. Teine lause tähendab, et politsei hoiab kuritegevust ohjes [2].

3. Selge keele liikumine

Kantseliidivaba ja selgema keelekasutuse poole püritakse praegusel ajal mitmel pool maailmas. Euroopas tegeleb selgema keele arendamisega ühendus Plain Language Europe [12], milles osaleb 6 riiki, kuid mitte veel Eesti. Arusaadavate ja konkreetsete tekstide loomisele pööratakse aina rohkem tähelepanu. Siinses peatükis kirjeldatu tugineb Katrin Halliku tekstile „Selge keele liikumisest maailmas“ [13].

Selge keel on keel, mis on keskmisele inimesele arusaadav. Selgest keelest on eemaldatud mittevajalikud sõnad ning keerukas stiil, et tekst muuta lugejasõbralikumaks ning arusaadavamaks. See aga ei tähenda, et selge keel oleks lihtsustatud keel. Peale selle nimetatakse selget keelt ka kantselseikeele ehk bürokraatliku keele vastandiks.

Tänapäeval on selge keele liikumine muutunud koordineeritud tegevusteks paljudes riikides üle maailma. Näiteks koostas Hollandi firma BureauTaal Euroopa keeleõppe raamdokumendi, mille põhjal tehti programm Texamen. Texameniga hinnati Hollandi põhiseadust kindlate tunnuste järgi – nominalisatsiooni, kantseliitliku keelekasutuse, sõnasageduse, lausetepikkuste jt. Saadi teada, et 50% Hollandi elanikest sai aru põhiseaduse tekstidest. Peale seda pandi kokku meeskond, kes kirjutaks Hollandi põhiseaduse ümber eesmärgiga, et 95% Hollandi elanikest saaks põhiseadusest aru. Eesmärk saavutati ning sellest ajast peale kasutatakse Hollandis ümberkirjutatud põhiseadust paralleelselt algse põhiseadusega.

Ka selge keele akadeemilise koolituse vajadus on Euroopas suurenenud. 2010. aastal loodi rahvusvaheline projekt „Selge keel“. Selle projekti lõi Belgia Katholieke Hogeschooli õppejõud ja nende eesmärk oli koostada selge keelekasutuse ja infodisaini õppekava ning rakendada seda Euroopa ülikoolides üheaastase diplomiõppe kursusena.

1983. aastal Inglismaal algatati vabatahtlik ühendus Clarity. Clarity on rahvusvaheline organisatsioon, mis koosneb juristidest, keeleinimestest ja ametnikest. Clarity-l on praegu üle 1000 liikme enam kui 50 riigis. Clarity eesmärk on propageerida selget keelt seadusloomes, vältida arhailist, ebaselget ja keerukat juriidilist keelt.

4. Sarnased programmid

Eestikeelsete tekstide jaoks ei ole programmi, mis tuvastaks tekstist kantseliiti ja paronüüme. Samuti pole teada ka sellist funktsiooni mõne teise programmi koosseisus. Mõnede suuremate keelte tekstidele on selliseid rakendusi siiski arendatud. Allpool kirjeldatakse mõningaid selliseid programme.

4.1 Microsoft Word

See alapeatükk on refereeritud kasutades Microsoft Wordi ametlikku dokumentatsiooni. [14]

Microsoft Wordi rakendus suudab eestikeelsete tekstide jaoks küll õigekirja parandada, kuid ingliskeelse teksti jaoks on teda rohkem arendatud. Näiteks saab Microsoft Wordi abil leida ingliskeelsest tekstist üles keerukamad sõnad ja seejärel soovitab programm kasutajale mõnda lihtsamat, kuid tähenduselt sama sõna. Järgnevalt paar näitlikku lausepaari kantseliidi kohta:

- 1) *The magnitude of the problem is far beyond the scope of humanitarian aid.* (tõlge eesti keelde – „*Probleemi magnituud on kaugel humanitaarabi ulatusest*“).
- 2) *The size of the problem is far beyond the scope of humanitarian aid.* (e.k – „*Probleemi suurus on kaugel humanitaarabi ulatusest*“).

Esimene näitelause on Microsoft Wordi sisestatud tekst ning teine lause on parandatud tekst. Microsoft Word muudab esimeses lauses sõna „*magnitude*“ (e.k – „*magnituud*“) sõnaga „*size*“ (e.k – „*suurus*“). See muudab teksti arusaadavamaks ning annab sõnumi selgemalt edasi.

Ingliskeelsest tekstis suudab Microsoft Word leida üles ka nominaalstiili, mis on kantseliitlikule tekstile omane stiil. Microsoft Word proovib teksti muuta aktiivsemaks, eemaldades üleliigsed nimisõnad ning lisades nende asemele tegusõnu. Näiteks:

- 1) „*The trade union is holding negotiations with the employers.*“ (e.k – „*ametühing peab läbirääkimisi töötajatega*“).
- 2) „*The trade union is negotiating with the employers.*“ (e.k. – „*ametühing räägib töötajatega läbi*“).

Toodud näites muudab Microsoft Word tegusõna ja nimisõna paari „*holding negotiations*“ tegusõnaks „*negotiating*“, mille tagajärjel muutub tekst konkreetsemaks ja lühemaks.

Microsoft Word on suuteline leidma ingliskeelsest tekstist üles ka liigset informatsiooni, näiteks:

- 1) „*Her backpack was large in size.*“ (toortõlge e. k. – „*Tema seljakott oli suuruselt suur*“).
- 2) „*Her backpack was large.*“ (toortõlge e. k. – „*Tema seljakott oli suur*“.)

Selles näites on esimeses lauses sõnu, mis ei lisa lausele teadmist. Microsoft Word on võimeline sellest aru saama ning eemaldab sõnapaari „*in size*“.

Peale kantseliidi leidmise suudab Microsoft Word leida ingliskeelsest tekstist üles ka paronüüme, näiteks:

- 1) „*Could you please advice me?*“ (toortõlge e.k. – „*Kas te palun nõu mulle?*“)
- 2) „*Could you please advise me?*“ (toortõlge e.k. – „*Kas annaksite mulle nõu?*“)

Siin on tegemist inglise keeles leiduva paronüümpaariga – „*advice*“ ja „*advise*“. Sõna „*advice*“ on nimisõna – eesti keeles „*nõuanne*“ ning sõna „*advise*“ on tegusõna – eesti keeles „*nõu andma*“. Microsoft Word muutis sõna „*advice*“ sõnaks „*advise*“.

4.2 OpenOffice – LanguageTool

LanguageTool on programm, mis on suuteline leidma tekstist nii kantseliiti kui ka paronüüme. Sellest on saadaval nii veebipõhine kui ka allalaetav versioon. Samuti on see integreeritav erinevatesse veebibrauseritesse, Google Docsi keskkonda ning LibreOffice programmidesse. LanguageTool toetab enam kui 20 keelt, kuid mitte eesti keelt [15]. Selle programmi testimisel sisestati samad lausetepaarid eelmisest peatükist.

„*Could you please advice me?*“ puhul märgendas LanguageTool sõna „*advice*“ ning selle sõna peale klikkides tuli aken, kus soovitati lausesse sobivamat sõna – „*advise*“.

„*Her backpack was large in size.*“ Puhul märgendas programm lause osa „*large in size*“ ning sellele märgendile peale vajutades soovitas programm jätta alles ainult sõna „*large*“.

Rakendusega prooviti tekstist leida ka nominaalstiili parandusi. Sisestades lauset „*The trade union is holding negotiations with the employers.*“, ei leidnud LanguageTool tekstist ühtegi viga, kuid Microsoft Word tuvastas täpselt samas lauses vea. Lause „*The magnitude of the problem is far beyond the scope of humanitarian aid.*“ korral ei märkinud LanguageTool ühtegi sõna, Microsoft Word, aga muutis ühe sõna lihtsamaks.

5. Programmi ülevaade

Programmi tegemiseks kasutati programmeerimiskeelt *Python*. Keel valiti sellepärast, et selles keeles on olemas EstNLTK teek ning graafilise liidese loomine tkinteriga mooduliga on kergesti õpitav. Programmi andmebaasiks, mille järgi programm kantseliidi ja paronüümide üle otsustab, on kasutatud raamatuid „Väike paronüümisõnastik“ [2] ning „Kuidas hoiduda kantseliidist“ [1]. Esimesest saadi paronüümid ning teisest kantseliitlikud lausekonstruktsioonid ja väljendid, mida programm tuvastama hakkab.

Programmi käivitamise juhend asub lisas 1. Programmi kood on tööle lisatud tihendatud failina.

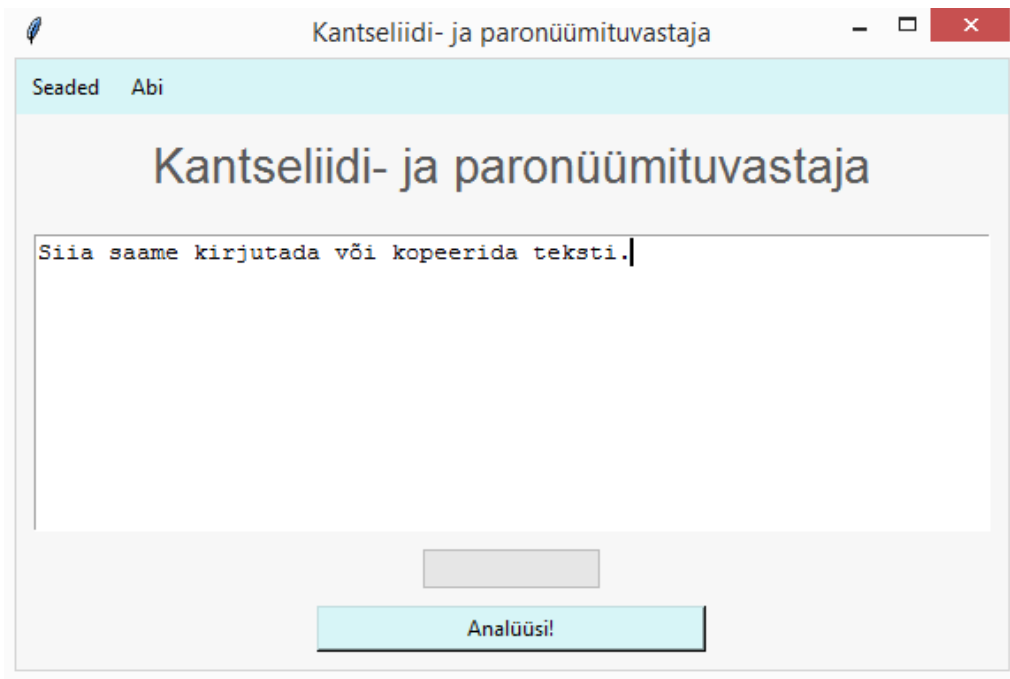
5.1 EstNLTK

Järgnev lõik tugineb EstNLTK ametlikule leheküljele [16]. EstNLTK (*Estonian natural language processing toolkit*) on programmeerimiskeele *Python* avaliku lähtekoodiga eesti keele keeletehnoloogilise töötluse teek. EstNLTKd kasutatakse näiteks teksti lõikudeks, lauseteks või sõnadeks jupitamiseks. EstNLTKs on olemas morfoloogiline analüsaator (sõnaliigi määramine, sõnavormi määramine jne), süntaktiline analüsaator, lemmatiseerija jpm. Siinses töös kasutati morfoloogilist analüüsi, süntaktilist analüüsi ning lemmatiseerimist. Morfoloogilise analüüsiga saadakse sõnade sõnavorme ja sõnaliike. Süntaktilise analüüsiga saadakse lauses olevate sõnade süntaktilised märgendid (nt sihitis, öeldistäide) ja sõnade süntaktilised ülemad süntaksipuus.

5.2 Tkinter

Tkinter [17] on *Pythonis* kõige populaarsem graafilise liidese teek. Tkinter on kergesti õpitav ning võimaldab teha kerge vaevaga lihtsaid kasutajaliideseid. Kantseliidi tuvastamisprogrammi kasutajaliidese jaoks on kasutatud just seda teeki.

5.3 Programmi kasutajaliides



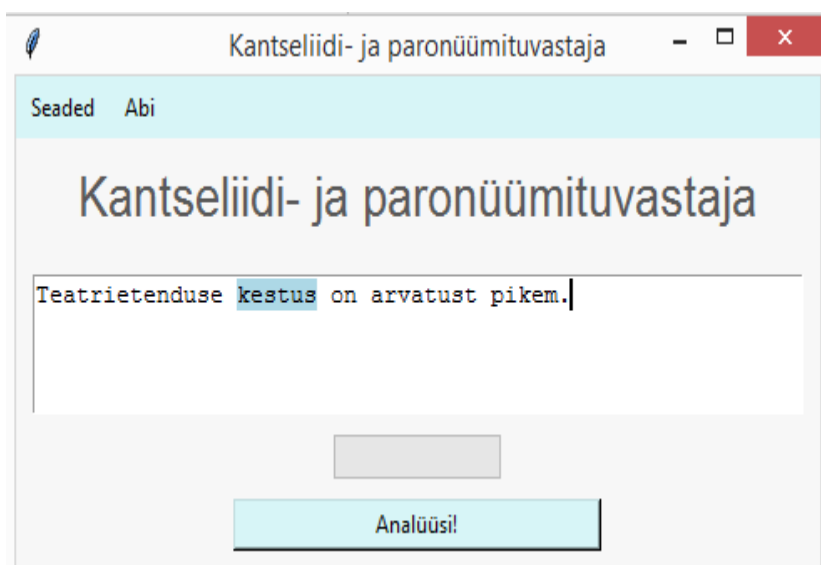
Joonis 1. Programmi kasutajaliides

Joonisel 1 on näha, milline loodud programm välja näeb. Programmi algne aken koosneb menüüribast, tekstikastist ning nupust „Analüüsi!“. Kui tekstikasti on kopeeritud või kirjutatud mingi tekst ning vajutatakse nuppu „Analüüsi!“, siis programm märgib tekstist leitud paronüümid ning kantseliidi. Menüüriba nupult „Seaded“ saab kasutaja valida, milliseid kantseliidi vorme programm otsib. Lisaks saab valida, kas programm otsib tekstist paronüüme. Menüüriba nupult „Abi“ saab tutvuda programmi kasutamishendiga ning vaadata statistikat – kui palju leiti tekstist mingit kantseliidi vormi ning mitu paronüümi kokku leiti. Nupu „Analüüsi!“ kohal on edenemisriba. Programm analüüsib lehekülje pikkust teksti umbes 4-5 sekundit.

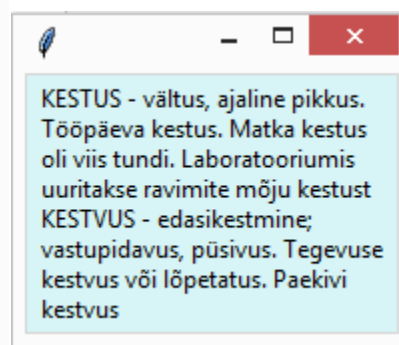
5.4 Paronüümide märkimine

Programm kasutab paronüümide tuvastamiseks tekstidokumenti, mis on sobivale kujule viidud „Väike paronüümisõnastik“. Paronüümide tekstidokumendis on paronüümid omavahel grupeeritud ning iga paronüümi järel on paronüümi seletus. Näiteks on üheks paronüümigrupiks „kestus“ ja „kestvus“. Mõlema sõna järel on sõnaseletus. Igas paronüümigrupis on kaks või enam paronüümi. Dokumendis on kõik paronüümid esitatud sõna algvormi ehk lemmana. Programm suudab tuvastada kokku 323 paronüümi.

Tekstiaknasse sisestatud tekstist paronüümide leidmiseks kasutab programm EstNLTK lemmatiseerijat, mis kontrollib iga sisendtekstis olevat sõna ning viib sõna algvormile. Seejärel otsitakse, kas mõni sisendtekstist leitud lemma on olemas ka paronüümide tekstidokumendis. Kui lemma leitakse, siis märgendatakse see sõna programmi tekstikastis. Paronüümid on tähistatud sinise värviga.



Joonis 2. Paronüümide märkimine



Joonis 3. Paronüümi infoaken

Joonisel 2 on tekstikasti sisestatud lause „Teatrietenduse kestus on arvatust pikem”. Kuna sõna „kestus“ lemma leidis paronüümide tekstidokumendis, siis see sõna märgendati. Kui märgendile peale vajutada, siis avaneb infoaken (joonis 3), kus on paronüümigrupi sõnade „kestus“ ja „kestvus“ seletused.

5.5 Kantseliidi märkimine

Kantseliidi tuvastamiseks kasutab programm andmeid ja kantseliidi määratlusi, mis on saadud raamatust „Kuidas hoiduda kantseliidist“ [1]. Selles raamatus on neli peatükki, mille igas peatükis käsitletakse erinevaid kantseliitlikke vorme. Loodud programm katab neist enamiku. Esimese peatüki kantseliitlikud vormid jäid programmi tegemisel välja, sest nende kaasamine osutus antud töö raames liiga mahukaks. Näiteks peatükis 1.4 käsitletakse sõnavalikut, mille automaatne hindamine vajaks semantilist analüüsi. Samuti jäid programmist välja teise peatüki konstruktsioonid, mida käsitleti peatükkide „parem on *da*-tegevusnimi“ ja „parem on *des*-vorm“ all.

Trükise „Kuidas hoiduda kantseliidist“ teisest peatükist hõlmab programm nominalisatsiooni *mine*-vorme koos tühisõnadega, *poolt*-tarindit ning konstruktsiooni *olema + kesksõna*. Kolmas peatükk koosneb kantseliitlikest sõnadest ning neljas peatükk koosneb kantseliitlikest vormidest: *It*-määrsõnad, määrus saavas käändes ning mitmus ainsuse asemel. Kolmas ja neljas peatükk on programmi täielikult kaasatud.

5.6 Kantseliidi märkimine kasutades lemmatiseerimist ja süntaktilist analüüsi

Programm kasutab sõnade ning vormide märkimisel põhiliselt kahte EstNLTK funktsiooni. Üks on sõnade märkimine kasutades lemmatiseerijat ning teine süntaktiline analüüs. Süntaktilise analüüsiga saame lause sõnade omavahelised sõltuvussuhted. Töös on süntaksi märgendamiseks kasutatud EstNLTK MaltParserit [16].

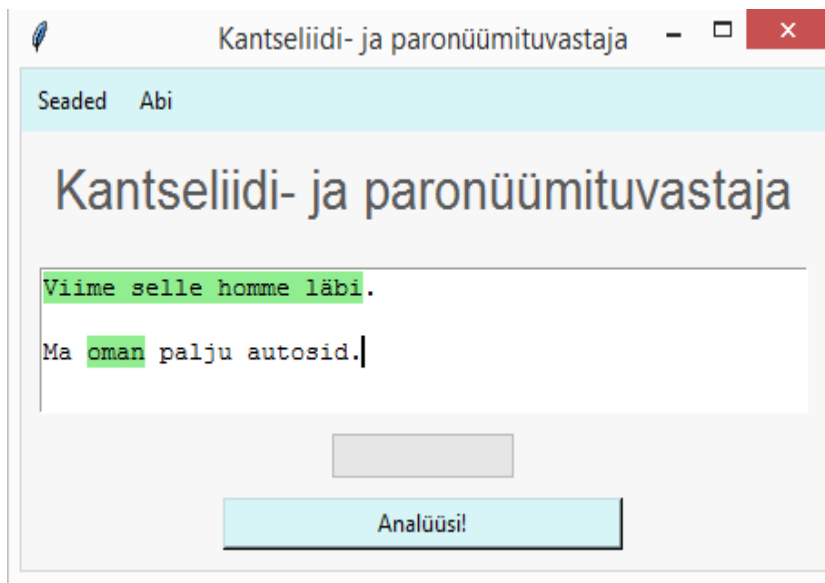
5.6.1 Tarbetud sõnad

Tarbetute sõnade kirjeldus on saadud raamatust „Kuidas hoiduda kantseliidist“ [1]. Raamatus kirjutatakse, et tarbetute sõnade asemel saab tihti kasutada lühemaid ja selgemaid sõnu. Tarbetud sõnad hõlmavad endas kantseliitlikeid viitefraase (nt *eelnev*, *eelmine*, *järgmine* jne), nugisõnalisi üldlaiendeid (nt *praktiliselt*, *reeglina*), võõrsõnu, mis teevad teksti lugemise keerulisemaks ning muid tarbetuid vorme (nt *vastavalt sellele*, *alljärgnevalt*, *käesolevaga tõendan*).

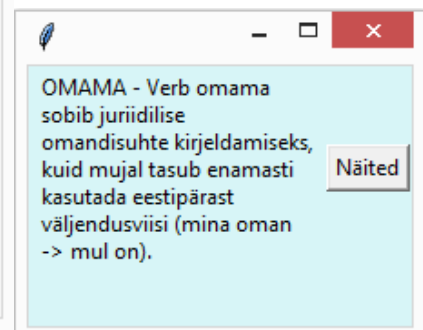
Programm kasutab tarbetute sõnade leidmisel raamatu „Kuidas hoiduda kantseliidist“ [1] kolmandat peatükki. Selles peatükis esinevad sõnad sõnaliikide kaupa: tegusõnad, nimisõnad, omadussõnad, kaassõnad ning määrsõnad. Iga sõna jaoks on olemas sõna seletus ning näitelauseid – kuidas kindlast kantseliitlikkust sõnast hoiduda. Kõik peatükis olevad sõnad on viidud programmi jaoks tekstidokumendi koos sõnade seletuste ja näitelauseetega. Osad neist sõnadest on algvormina, aga mitte kõik. Lisaks pole alati tegu ainult üksikute sõnadega, vaid nende hulgas leidub ka ühendverbe (nt *läbi viima*, *paika panema*). Kantseliitlike tarbetuid sõnu, mida programm otsib, on kokku 28.

Tarbetute sõnade leidmiseks käib programm kõik sisendteksti sõnad läbi. Kuna eelpool mainitud tekstidokumendis ei ole kõik tarbetud sõnad lemmadena, siis kõigepealt otsitakse sõnu tekstidokumendist lemmatiseerijat kasutamata. Kui sõna tekstidokumendist ei leita, siis võetakse sõna lemma ning otsitakse uuesti. Sõna leidmisel märgitakse see programmi tekstikastis roheline värviga.

Kuna tekstidokumendis leidub ka ühendverbe ja ühendverbi sõnad ei esine lauses alati järjest (nt ühendverb „läbi viima“ võib esineda lauses kujul „viisime selle läbi“, kus sõna „selle“ on ühendverbi vahel), siis selle probleemi lahendamiseks kasutati süntaktilist analüüsi. Süntaktilist analüüsi tehes käiakse kogu sisendtekst lausete kaupa läbi. Kui lausest leitakse ühendverbi esimene sõna ning kui selle sõna ülemus on sama ühendverbi teine sõna (verb), siis on need sõnad omavahel süntaktilises seoses. Kui selline ühendverb leitakse, märgitakse mõlemad sõnad ning kogu tekst, mis nende kahe sõna vahele jäävad. Märgitud sõna on tähistatud rohelise värviga.



Joonis 4. Tarbetute sõnade märkimine



Joonis 5. Tarbetu sõna infoaken

Joonisel 4 on programmi sisestatud kaks lauset „Viime selle homme läbi“ ja „Ma oman palju autosid“. Esimeses lauses on märgitud roheliseks kogu lause, sest sealt leiti sõna „läbi“, mis on süntaktilises seoses sõnaga „viime“. „Viime“ lemma on „viima“, järelikult on tegemist ühendverbiga „läbi viima“. Teises lauses leiti sõna „oman“, mille lemma „omama“ on tarbetute sõnade tekstidokumendis. Ka see sõna märgiti rohelisega. Joonisel 5 on näha infokasti, mis avaneb, kui vajutada sõna „oman“ peale.

5.6.2 Nominalisatsiooni *mine*-vormid

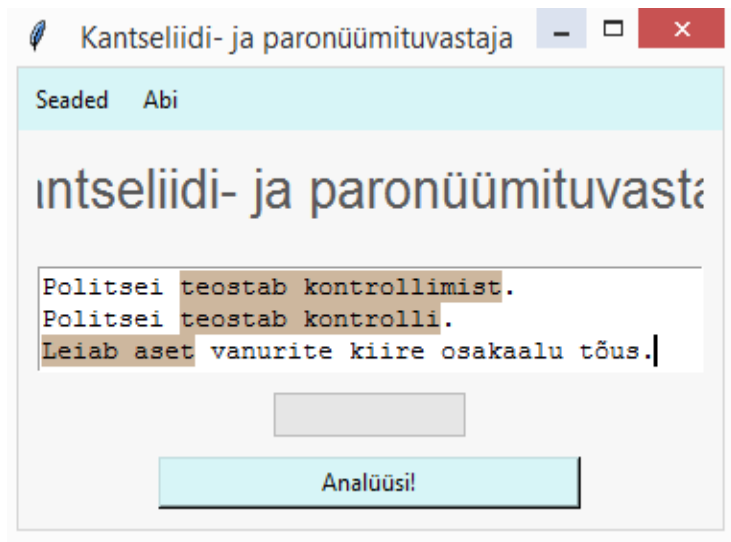
Nimisõnatüveks ehk nominalisatsiooni liigtarvitamiseks peetakse *mine*-vormi kasutust koos tühi- ehk tugiverbiga (nt *teostama*, *kuuluma*). See lisab tekstile tegusõnadest moodustatud nimisõnu, mis muudavad teksti abstraktsemaks [1]. Sagedasemad tühiverbid, millega nominalisatsiooni

tekitatakse, on *teostama*, *toimuma*, *kuuluma*, *vajama* ja *aset leidma*. Näiteks sõnaühendit „*teostame kontrollimist*“ saame öelda ühe sõnaga „*kontrollime*“.

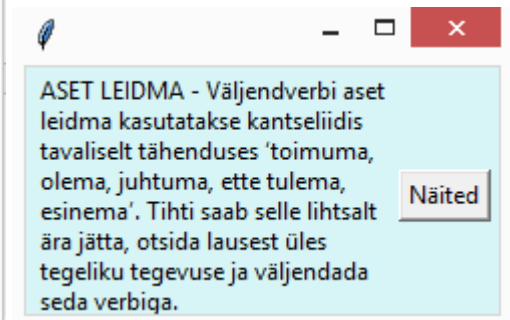
Programm kasutab nominalisatsiooni *mine*-vormide leidmiseks informatsiooni raamatust „Kuidas hoiduda kantseliidist“ [1] peatükist 2.1. Raamatus on 5 tühiverbi, igal verbil on juures seletus ning näitelause, kuidas neist hoiduda. Kõik see informatsioon on programmi jaoks viidud tekstidokumendi. Dokumendis olevad tühverbid on sõnade algvormid.

Programm leiab tekstist nominalisatsiooni käies läbi kõik sisendteksti laused. Nominalisatsiooni otsitakse süntaktilise analüüsi abil, kasutades kahte lähenemist. Esimesena vaadatakse, kas lauses leidub sõna, mille lemma on *mine*-lõpuline (näiteks „*kontrollimine*“). Kui selline lemma leidub, siis vaadatakse ülemust süntaksipuus. Kui seostuva sõna lemma on tühverbide tekstidokumendis (näiteks „*teostama*“), siis märgitakse kogu fraas. Teisena vaadatakse, kas lauses leidub sõna, mille süntaksimärgend on sihitis ja sõnaliik nimisõna (näiteks „*kontrolli*“). Kui selline sõna leidub, vaadatakse tema süntaktilist ülemust. Kui ülemuse lemma on tühverbide tekstidokumendis, märgitakse kogu fraas.

Kuna tühverbide hulgas leidus ka ühendverb „*aset leidma*“ ning selle süntaktiline analüüs ei anna õigeid tulemusi, otsustati nominalisatsiooni ühendverbide jaoks kasutada sõnaakent +/- 4 sõna. Süntaktilise analüüsi korral tekib probleem, et ühendverbi sõnad „*aset*“ ja „*leidma*“ ei ole süntaktilises seoses. Näiteks lause „*Leiab aset vanurite kiire osakaalu tõus*“ korral on sõna „*tõus*“ ülemus sõna „*leiab*“, aga sõna „*leiab*“ ei ole kuulu ühendverbi sõna „*aset*“ juurde, vaid sõna „*vanurite*“ juurde. Seega kasutatakse sõnaakent, mis otsib lausest sõna „*aset*“ ning sellest kuni nelja sõna kauguselt sõna, mille lemma on „*leidma*“. Kui selline sõnaaken leitakse, märgitakse programmis kogu sõnaaken.



Joonis 6. Nominalisatsiooni märkimine



Joonis 7. Nominalisatsiooni Infoaken

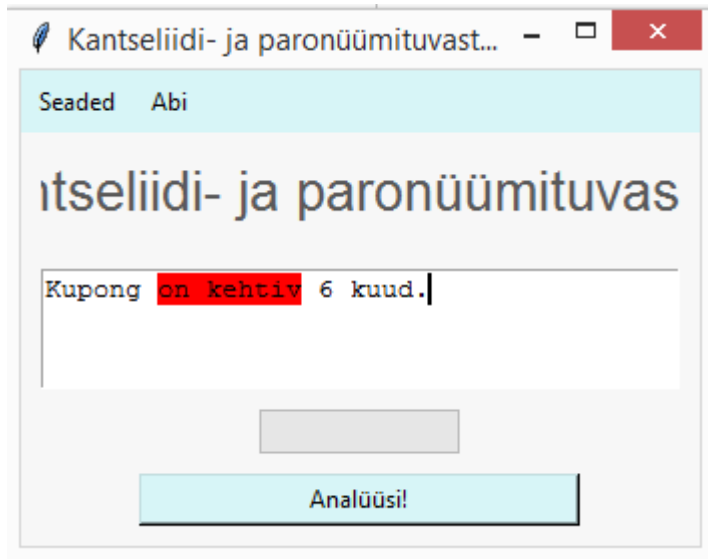
Joonisel 6 on sisestatud kolm lauset. Lauses „*Politsei teostab kontrollimist*” leiti sõna „*kontrollimist*“, mille lemma on *mine*-lõpuline. Sõna süntaktiline ülemus on sõna „*teostab*“, mille lemma on „*teostama*“. Teise lause korral leiti sõna „*kontrolli*“, mis on sihtis ja nimisõna. Süntaktiliselt on tema ülemussõna samuti „*teostab*“. Kolmanda lause korral kasutati sõnaakent. Leiti sõna „*aset*“ ning nelja sõna kauguselt leiti sõna, mille lemma on „*leidma*“. Märkend on pruuni värvi. Joonisel 7 on näha infokasti, mis avanes „*aset leidma*“ märgendile vajutades.

5.6.3 Olema + kesksõna

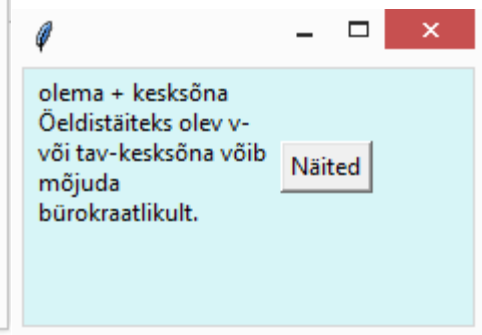
Öeldistäiteks olev *v*- või *tav*- kesksõna võib mõjuda bürookraatlikult, näiteks „*Pakkumine on kehtiv 6 kuud*“ saab paremini öelda „*Pakkumine kehtib kuus kuud*“ [1].

Programmi jaoks vajalik informatsioon *olema + kesksõna* kohta on saadud raamatu „*Kuidas hoiduda kantseliidist*“ peatükist 2.3 [1]. Raamatus on esitatud näitelauseid ning seletus, miks *olema + kesksõna* vorm halb on. See informatsioon on viidud programmi jaoks tekstidokumenti.

Olema + kesksõna otsimiseks käib programm kõik sisendteksti laused läbi. Kui lausest leitakse sõna, mis on öeldistäide ja mille lemma lõpus on *-v* või *-tav*, vaadatakse selle süntaktilist ülemat. Kui ülema lemma on „*olema*“, siis on tegemist *olema + kesksõna* vormiga. Programmi tekstikastis märgitakse nii öeldistäide, „*olema*“ vorm kui ka nende vahelised sõnad. Märkend on punast värvi.



Joonis 8. Olema + kesksõna märkimine



Joonis 9. Olema + kesksõna märkimine

Joonisel 8 on programmi sisestatud lause „Kupong on kehtiv 6 kuud“. Rakendus märkis sõnaühendi „on kehtiv“. Rakendus leidis tekstist sõna „kehtiv“, mis on predikaat ning tema lemma lõppeb v-ga. Sõna „kehtiv“ ülemus süntaksipuu on sõna „on“, mille lemma on „olema“. Joonisel 9 on infoaken, mis avaneb märgendile vajutades.

5.7 Kantseliidi märkimine kasutades morfoloogilist ja süntaktilist analüüsi

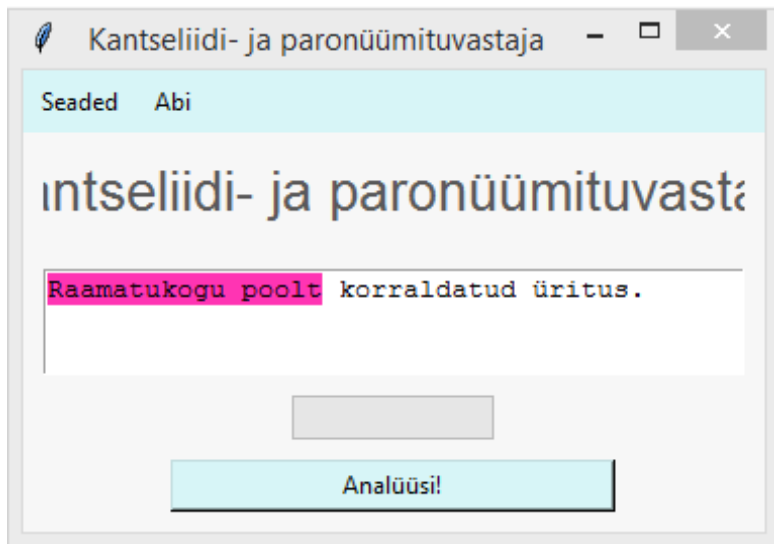
Selles peatükis kirjeldatakse, kuidas rakendus märgib kantseliiti kasutades morfoloogilist analüüsi (kasutatakse kas morfoloogilisi vorme või sõnaliike) või süntaktilist analüüsi ja morfoloogilist analüüsi.

5.7.1 Poolt-tarind

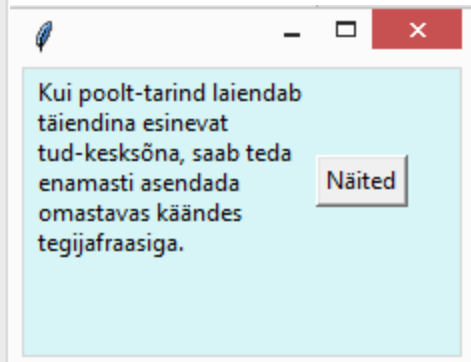
Poolt-tarind on võõrapärane vorm, mis eestikeelses lauses jätab tegija tagaplaanile. Näiteks lause „Operatsioon oli arsti poolt tehtud“ on kantseliitlik, aga selle lause asemel on võimalik kasutada „Arst tegi operatsiooni“.

Programm kasutab *poolt*-tarindi leidmisel raamatu „Kuidas hoiduda kantseliidist“ peatükis 2.2 olevaid näitelauseid. Näitelauseid on viidud programmi jaoks tekstifaili. *Poolt*-tarindi leidmisel kasutatakse ainult morfoloogilist analüüsi.

Poolt-tarindite otsimisel käiakse sisendteksti laused läbi. *Poolt*-mall märgitakse siis, kui lausest leitakse omastavas käändes sõna, millele järgneb sõna „*poolt*“. Märgitakse omastavas käändes sõna ja sõna „*poolt*“. Märgend on roosat värvi.



Joonis 10. *poolt*-tarindi märkimine



Joonis 11. *Poolt*-tarindi infoaken

Joonisel 10 on tekstikasti sisestatud lause „*Raamatukogu poolt korraldatud üritus*“. Lauses on märgitud kantseliidina *poolt*-tarind. Sõna „*raamatukogu*“ on omastavas käändes ning sellele järgneb sõna „*poolt*“. Joonisel 11 on infoaken, mis avaneb roosale märgendile vajutades.

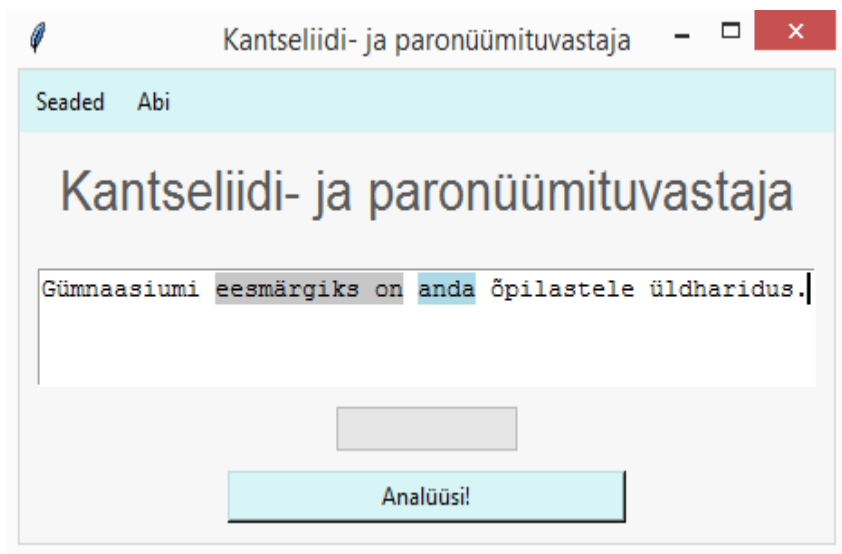
5.7.2 Määrus saavas käändes

Saavas käändes määrus märgib omaduste või seisundite juhuslikkust, ajutist iseloomu. Parem on kasutada kindlamat kõneviisi ja loobuda juhuslikkusest [1]. Näiteks lauseosa „*töö eesmärgiks on*“ oleks parem kirjutada „*töö eesmärk on*“.

Informatsioon selle kantseliidivormi kohta on saadud raamatu „Kuidas hoiduda kantseliidist“ peatükist 4.2 [1]. Raamatus selgitatakse, miks see kantseliidivorm halb on, ning esitatakse näitelaused selle kohta, kuidas vormi vältida. Seletus ja näitelaused on programmi jaoks viidud tekstidokumenti. Selle konstruktsiooni otsimisel kasutatakse nii morfoloogilist kui ka süntaktilist analüüsi.

Saavas käändes määruse otsimiseks käiakse läbi kõik sisendteksti laused. Kui lausest leitakse sõna, mille süntaksimärgend on määrus ja ta on saavas käändes, vaadatakse selle sõna süntaktilist

ülemust. Kui ülemuse lemma on „olema“, siis märgitakse mõlemad sõnad ja nende vaheline tekst kantseliidina.



Joonis. 12. Saavas käändes määruse märkimine

Joonisel 12 on tekstikasti sisestatud lause „Gümnaasiumi eesmärgiks on anda õpilastele üldharidus“. Rakendus märkis kantseliidina sõnaühendi „eesmärgiks on“, sest leiti, et sõna „eesmärgiks“ on määrus ja ta on saavas käändes. Kuna sõna „on“ lemma on „olema“ ja ta süntaktiline ülemus on sõna „eesmärgiks“, siis märgiti „eesmärgiks on“ kantseliidina. Kantseliidivorm määrus saavas käändes märgitakse halli värviga. Hallile märgendile peale vajutades avaneb infoaken, mis on sarnane joonisel 11 oleva infoaknaga, aga kantseliidi konstruktsiooni määrus saavas käändes kirjeldava infoga. Sinine märgend joonisel 12 näitab paronüümi.

5.7.3 *lt*-määrsõnad

Raamatust „Kuidas hoiduda kantseliidist“ [1] on tähelepanu all ka *lt*-määrsõnad, sest ületuletuse teel tekkinud *lt*-määrsõnad muudavad teksti keeruliseks. Näiteks on ületuletuse teel tekkinud sõna „koheselt“. Määrsõna „koheselt“ saadakse määrsõnast „kohe“. Sõnast „kohe“ on tuletatud omadussõna „kohene“ ning sellest omakorda määrsõna „koheselt“. Tegelikult tähendavad „koheselt“ ja „kohe“ täpselt sama asja. Ületuletuse saab tekitada ka sõnaühendist. Kui sõnaühendist „veel kord“ teha sõna „veelkordne“ ja sellest omakorda sõna „veelkordselt“, ka siis on tegemist kantseliitliku konstruktsiooniga.

Eespool osutatud raamatust on programmi jaoks saadud *lt*-määrsõnade seletus ning näitelauseid, kus *lt*-määrsõnu valesti kasutatakse ning kuidas *lt*-määrsõnu vältida. Näitelauseid ja seletus on viidud tekstidokumenti. Kantseliidi leidmiseks kasutab programm morfoloogilist analüüsi.

Lt-määrsõnade leidmiseks käiakse tekst sõna kaupa läbi. Ületuletuse teel tekkinud määrsõnade leidmiseks otsitakse sõnu, mis on *-selt*-lõpulised. Neilt võetakse *-selt* lõpust ära ning kui allesjäänud sõna on määrsõna, siis märgitakse sõna ära, nt *lt*-määrsõnadena märgitud ka sõnad *veelkordselt*, *esmakordselt*, *ööpäevaringselt* ja *igakuiselt*. *-selt*-liite eemaldamine nende sõnade peal ei toimi. Näiteks kui sõnalt *esmakordselt* võetakse *-selt* lõpust, siis alles jäänud sõna analüüsitakse kui nimisõna, mitte määrsõna. Eelnevate liitsõnade kontrollimiseks eemaldatakse tuletuse aluseks olnud omadussõna saamiseks sõna lõpust *-lt*. Siis eraldatakse liitsõna eraldiseisvateks lemmadeks ning kui liitsõna viimane sõna on kas *-kuine*, *-ringne* või *-kordne*, märgitakse kogu sõna kantseliidina. Märgendi peale vajutades avaneb infoaken, mis annab informatsiooni ning näitelauseid *lt*-määrsõnade kohta. Märgend on kollast värvi.

5.7.4 Mitmus ainsuse asemel

Eesti keeles on hakatud ainsussõnu kasutama mitmuses. Selline vale sõnade kasutamise viis on tulnud võõrkeeltest. Näiteks lause „*Olulised arengud on toimunud*“ asemel on parem kasutada ainsuslikku lauset „*Oluline areng on toimunud*“.

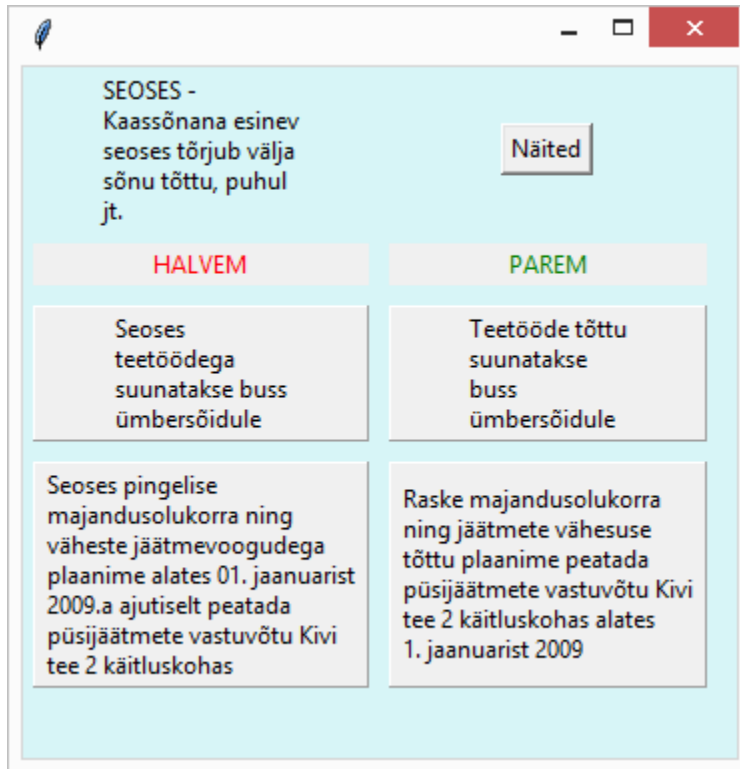
Informatsioon kantseliitliku mitmuse kasutamise kohta on saadud juba eelpool viidatud raamatu peatükist 4.3 [1]. Seal selgitatakse üldisemalt mitmuse väärkasutust, nt tuuakse välja kahe sõna „*arengud*“ ja „*tegevused*“ kasutus ning esitatakse kummagi sõna kohta näitelauseid. Sõnad, seletus ja näited on viidud programmi jaoks tekstidokumenti. Kantseliidi leidmiseks kasutatakse morfoloogilist analüüsi.

Selle kantseliidivormi korral käiakse kogu sisendtekst sõna kaupa läbi. Kui tekstist leitakse sõna, mille lemma on võrdne tekstidokumendis oleva sõna lemmaga, ning leitud sõna on morfoloogilise analüüsi kohaselt mitmuses, siis sõna märgendatakse. Märgendi peale vajutades avaneb infoaken, mis annab informatsiooni selle kantseliidivormi kohta. Märgend on tumeroosat värvi.

5.8 Infoaken

Nii paronüümi- kui ka kõikide kantseliidimärgendite peale saab vajutada ja see avab infoakna. Paronüümide korral antakse infoaknas tuvastatud sõnaga samas paronüümide grupis olevate

paronüümide kohta informatsiooni. Ka kantseliidi korral antakse infoaknas kindla kantseliidi kohta informatsiooni, kuid kantseliidi korral on infoaknal ka nupp „Näited“. Kui kasutaja vajutab nupule „Näited“, siis infoaken laieneb ning kantseliidi kohta näidatakse näitelauseid.



Joonis. 13. Infoaken

Joonisel 13 on näha laiendatud infoakent. Tekstikastist leiti kantseliitlik sõna „seoses“ ja kuvati selle märgendus. Märgendatud tekstikastile vajutamine avas infoakna. Joonis 13 illustreerib, milline näeb välja infoaken pärast nupule „Näited“ vajutamist. Laienenud infoaknal on kaks tulpa: „Parem“ ja „Halvem“. Vasakul pool tulbas on kantseliitlikud näitelauseid ning paremal pool tulbas on samad laused ilma kantseliitliku konstruktsioonita.

6. Edasiarendamise võimalused

Loodud programmi arendamiseks ja täiendamiseks on mitmeid võimalusi.

Esiteks saab programmi muuta kasutajasõbralikumaks, kui teha see veebis kasutatavaks, et tekiks võimalus midagi alla tõmbamata veebis oma teksti kontrollida või integreerida programm tekstiredaktorisse.

Teiseks saab programmi andmebaasi laiendada. Eesti keeles leiduvad kantseliitlikud vormid ei piirdu ainult raamatus „Kuidas hoiduda kantseliidist“ [1] olevate kantseliitlike konstruktsioonidega. Programmi saab arendada, et ta leiaks rohkem kantseliitlike konstruktsioone. Ka paronüüme on eesti keeles rohkem kui on toodud raamatus „Väike paronüümisõnastik“ [2]. Paronüümide andmebaasi saab suurendada. Võimalik oleks kasutada Eesti Keele Instituudi vastavaid andmebaase, mida pidevalt täiendatakse.

Kolmandaks saab programmile lisada semantilise analüüsi võimaluse, mis teeks süntaktilise analüüsi täpsemaks.

Neljandaks saab programmi arendada nii, et ta arvestaks konkreetse lausega, kontrolliks ka tuvastatud sõnade konteksti ja pakuks kasutajale leitud vigade parandamise võimalusi. Valminud töös kontrollitakse sõna esinemist, kuid ei kontrollita, kas paronüümi on kasutatud õigesti või valesti. Kui näiteks programm saab aru, et sõna „kestus“ on antud kontekstis valesti kasutatud, võiks ta pakkuda lause parandamiseks välja sõna „kestvus“.

Kokkuvõte

Bakalaureusetöö eesmärk oli luua programm, mis suudab eestikeelsest tekstist leida ning märkida paronüüme ja kantseliiti. Loodud programm tugineb paronüümide leidmisel raamatu „Väike paronüümisõnastik“ [2] ning kantseliidi leidmisel raamatu „Kuidas hoiduda kantseliidist“ [1] andmetele.

Töö eesmärk sai täidetud. Töö käigus loodi programm, mis leiab kõik raamatus „Väike paronüümisõnastik“ esitatud paronüümid ning enamiku kantseliitlikest konstruktsioonidest raamatust „Kuidas hoiduda kantseliidist“. Programm kasutab kantseliidi ja paronüümide tuvastamiseks tekstist morfoloogist ja süntaktilist analüsaatorit. Leitud paronüümid märgitakse siniselt ning erinevad kantseliitlikud vormid erinevate värvidega. Programm annab märgendatud konstruktsioonide ning sõnade kohta õige sõna valimiseks vajalikku informatsiooni ning kantseliidi korral esitab ka näitelauseid. Lisaks on programmil funktsioon, millega on võimalik ühe või mitme kantseliidivormi või paronüümide otsimine välja lülitada. Samuti on võimalik vaadata teksti statistikat, mis näitab, kui palju leiti tekstist üht või teist tüüpi kantseliidivorme ja/või paronüüme.

Kuna maailmas on hakatud rohkem tähelepanu pöörama puhta ja selge teksti kirjutamisele ning algatatud selge keele liikumine, mille eesmärk on muuta tekstid lugejasõbralikumaks ning arusaadavamaks, siis tuleks ka eesti keelel põhinevat programmi edasi arendada.

Rakendust on võimalik arendada seda näiteks veebikeskkonda viies. Sellisel juhul oleks ta paremini kättesaadav ning mugav kasutada. Teiseks on võimalik programmi lisada veel suuremal hulgal erinevaid kantseliitlike konstruktsioone ning laiendada paronüümide andmebaasi (nt sidudes andmebaasi Eesti Keele Instituudi pidevalt täiendatava andmebaasiga). Kolmandaks võiks programm kasutada semantilist analüsaatorit, sest sellisel juhul oleks kantseliidi märkimine täpsem. Neljandaks võiks programm anda kasutajale ka soovitusi, kuidas konkreetset lauset paremaks ja kantseliidivabamaks teha või paronüümi õigesti kasutada.

Viidatud kirjandus

- [1] Pullerits E. Kuidas hoiduda kantseliidist. Tartu: Keelehooldekeskus. 2010.
- [2] Plado H., Mandra K. Väike paronüümisõnastik. Tartu: Keelehooldekeskus. 2008.
- [3] Vettik A. Mis on mis. *Õiguskeel*, 1996, nr 1, lk 32.
- [4] Paet T., Rehema T. *Õiguskeel*, 2014, nr 2.
- [5] Keelehooldekeskuse kodulehekül. <http://keelehooldekeskus.ee/> (11.05.2017).
- [6] Liivaku U. Kust king keelt pigistab. Tallinn: Valgus. 1972.
- [7] Ress K. *Et lugeja võiks tulla*. Tallinn: Eesti Keele Sihtasutus, 2010, lk 42–43.
- [8] Argus R. Kantseliit, mis lööb oimetuks. *Director*, 2014, nr 2, lk 36-37.
- [9] Pullerits E. Kuidas hoiduda kantseliidist. *Oma Keel*, 2011, nr 2, lk 91.
- [10] Ametniku keelekäsiraamat. Tallinn: Juura, Õigusteabe AS. 2000.
- [11] Pullerits E. Keelehooldebrošüürid õpetaja ja õpilase vaatevinklist. *Oma Keel*, 2010, nr 1, lk 99.
- [12] Plain Language Europe. <https://www.plainlanguageeurope.com/en/> (11.05.2017).
- [13] Hallik. K. Selge keele liikumisest maailmas. *Õiguskeel*, 2011, nr 3.
- [14] Select grammar and writing style options in Office 2016. <https://support.office.com/en-us/article/Select-grammar-and-writing-style-options-in-Office-2016-eed60e9f-6b2e-4070-b30c-42efa6cff55a> (11.05.2017).
- [15] LanguageTool kodulehekül. <https://www.languagetool.org/> (11.05.2017).
- [16] EstNLTK – Open source tools for Estonian natural language processing. <https://estnltk.github.io/estnltk/1.4.1/index.html> (11.05.2017).
- [17] *Pythoni* Tkinteri dokumentatsioon. <https://wiki.python.org/moin/TkInter> (11.05.2017).

Lisad

Lisa 1. Programmi käivitamisjuhend

Programmi on proovitud Pythoni versiooniga 3.5.2 ja EstNLTK versiooniga 1.4.1

Juhend programmi käivitamiseks:

- 1) Tõmmata alla Pythoni 3.5.2 versioon
- 2) Installeerida EstNLTK viimane versioon
- 3) Kontrollida, et zip-failis olev kaust ja pythoni failid oleksid ühes kaustas
- 4) Jooksutada faili tuvastaja.py

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kaarel Sõrmus,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose Kantseliidi- ja paronüümituvastaja, mille juhendaja on Krista Liin.
 - 1.1.reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
 - 1.2.üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, **11.05.2017**