

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Ruth Schihalejev

Reeglipõhine lausekonstruktsioonide lihtsustaja eestikeelsetes tekstides

Bakalaureusetöö (9 EAP)

Juhendajad Sven Aller

Heili Orav

Tartu 2019

Reeglipõhine lausekonstruktsioonide lihtsustaja eestikeelsetes tekstides

Lühikokkuvõte

Selles bakalaureusetöös tutvustatakse praktilise osana tehtud keeleõppijatele mõeldud veebirakendust, mis lihtsustab eestikeelse teksti lausestruktuuri. Programmi eesmärk on teha liitlausest etteantud reeglite põhjal eesti keele grammatikale vastavad lihtlauseid. Rakenduse loomisel toetuti eesti keele analüüsimiseks mõeldud tehnilistele vahenditele, võtmetähtsusega on osalausestaja, kuid kasutati veel süntaksianalüsaatorit, morfoloogilist analüsaatorit ja morfoloogilist süntesaatorit.

Võtmesõnad:

Eesti keel, reeglipõhine lihtsustamine, süntakiline lihtsustamine, süntaks

CERCS: P175, Informaatika, süsteemiteooria

Rule-based Syntactic Simplifier for Texts in Estonian

Abstract

This Bachelor's thesis introduces an implemented web-based application that simplifies text in Estonian syntactically. The app is meant for language learners and its main propose is to make grammatically correct simple sentences from composite sentences by the given rules. This application relies on tools that are meant for text analysis in Estonian, mainly a clause segmenter, but syntactic and morphological analysers and a morphological synthesizer are used as well.

Keywords:

Estonian, Rule-based Simplification, Syntactic Simplification, Syntax

CERCS: P175, Informatics, systems theory

Sisukord

Sissejuhatus	5
1. Tekstide mõistetavaks tegemine	7
1.1. Teksti lihtsustamise viisid	7
1.1.1. Leksikaalne lihtsustamine	8
1.1.2. Süntaktiline lihtsustamine	8
1.1.3. Tekstist põhilise ekstraheerimine	9
1.2. Teksti automaatne lihtsustamine teiste keelte näitel	9
1.2.1. Uuringud süntaksi lihtsustamiseks	10
1.2.2. Süntaksi lihtsustamise rakendus YATS demo, Simplext	12
2. Eesti keele automaatseks analüüsiks loodud tehnilised vahendid.....	14
2.1. EstNLTK	14
2.2. Eesti keele süntaksianalüsaator	15
2.3. Eesti keele morfoloogiline analüsaator, ühestaja ja süntesaator	15
3. Eestikeelse teksti lauseehitust lihtsustav rakendus	17
3.1. Rakenduse prototüüp	17
3.2. Veebirakenduse ülevaade	18
3.3. Lihtsustamisprogrammi algoritm	19
3.3.1. Lause lihtsustamine	20
3.3.2. Jutumärke sisaldava lõigu analüüs	23
3.4. Rakendusele antud tagasiside	24
3.4.1. Hinnang muudetud lausetele	24
3.4.2. Hinnang muutmatuks jäänud lausetele.....	28
3.4.3. Kasutaja katsetuste tulemused.....	30
3.4.4. Kokkuvõtte küsitluse tulemustest.....	30

3.5. Probleemid.....	31
3.6. Edasiarendusvõimalused	34
Kokkuvõte	36
Viidatud kirjandus	38
Lisad	43
I. Küsitlus	43
II. Litsents	55

Sissejuhatus

Maailmas avaldatakse mitmel pool huvi Eesti riigi, keele ja kultuuri vastu, lisaks oleme tuntud oma infotehnoloogia kiire arengu poolest ning meil on palju e-residente. Eesti keelt õpetatakse välismaal umbes 60 üldhariduskoolis, lisaks saab kursusi võtta seltsidest ja ringidest („Eesti keele õpe...“, i.a.). Rohkem kui 30 välismaa kõrgkoolis õpib eesti keelt umbes 1000 inimest („Eesti keele välisõpe“, i.a.) ning Tartu Ülikooli eesti keele ajaloo ja murrete professori Karl Pajusalu suulise väite (2018) järgi on mõnes neist kõrgkoolidest üliõpilasi antud erialal rohkem, kui neid on Tartu Ülikoolis. Eestist saadetakse välismaale õpetajaid nii üheks kursuseks, terveks semestriks kui ka pikemaks perioodiks, kuid näiteks sel aastal on tulnud üliõpilased Hiinast, et tutvuda lähemalt oma erialavalikuga selle keskkonnas (Gao ja Shen, 2018). Ei tohi unustada keeleõppijaid ka Eesti Vabariigis: siin õpitakse eesti keelt nii esimese kui ka teise keelena. Ka meie pindalalt väike riik on tõmmatud globaliseerumise protsessi ning siia tuleb õppima ja töötama üsna palju inimesi välisriikidest.

Vaatamata eelarvamusele, et eesti keel on keerukas, üritatakse seda ikkagi omandada ning igapäevaste keelekasutajate võimuses on aidata nii keele õppijaid kui ka tarbijaid. Väga paljude (veebi)õpikute, -kursuste ja õppemängude kõrval jätab olukord automaatprogrammide abil keele kasutamist toetada soovida, sest igapäevased tarbetekstid on sageli väga keerulised. Tekstide lihtsamaks tegemiseks on varasemalt loodud bakalaureusetöö raames näiteks sõnavara lihtsustamise veebirakendus (Peedosk, 2017a) ja tekstide loojatele suunatud kantseliidi- ja paronüümituvastaja (Sõrmus, 2017).

Siinse bakalaureusetöö teema on teksti lausestuse lihtsustamine. Teemavalik on ajendatud isiklikust soovist ühendada peaeriala informaatika ja kõrvaleriala eesti keele teadmised ning aidata kaasa eesti keele püsijäämisele selle leviku kaudu. Töö käigus loodav rakendus võimaldaks eelkõige keele õppijatel lugeda ning mõista lauseehituselt keerulisemaid tekste. Samuti võib sellest abi olla õpikute tekstide lihtsustamisel (Mikk, 2015) või edaspidises tekstide keelelises masintöötuses (Chandrasekar, Doran ja Srinivas, 1996: 1041).

Töö eesmärk on luua programm, mis lihtsustab etteantud eestikeelse teksti lausestust, sh muutes liitlaused kõiki eestikeele grammatikaga nõutud osi sisaldavateks lihtlauseteks. Uurimistöös keskendutakse teksti erinevatele lihtsustamisvõimalustele ning sellele, milliseid süntaksi suunitlusega programme on teistele keeltele loodud. Ülesehituselt koosneb kirjalik osa kolmest sisupeatükist. Esimeses peatükis tutvustatakse teksti lihtsustamise viise

ning uuringuid süntaksi lihtsustamise kohta teistes keeltes. Teises antakse ülevaade eesti keele jaoks mõeldud tehnilistest vahenditest, mida koostatavas programmis kasutatakse. Kolmandas kirjeldatakse valminud veebirakendust, töötamise algoritmi ning analüüsitakse selle efektiivsust läbiviidud küsitluse põhjal ja tuuakse edasiarenduse võimalused.

1. Tekstide mõistetavaks tegemine

Keelt kasutavad paljud isesuguse tausta ja harjumustega inimesed. Peaaegu iga keele puhul on nii emakeele kui ka võõrkeelena kasutajaid ning erinevate tasemetega lugejaskondadele luuakse neile kohandatud tekste. Algtasemel rääkijate, näiteks väikelaste ja keeleõppijate, jaoks on vajalikud lihtsama keelega tekstid, kus esinevad pigem lühemad laused ja sagedasem sõnavara, kõrgema tasemega rääkijate puhul võib tekst sisaldada ka spetsiifilisemat sõnavara ning pikemaid lausekonstruktsioone (Rybing, Smith ja Silvervarg, 2010).

Lihtsama keelekasutusega seoses propageeritakse üle maailma selge keele¹ kasutust, millega suunatakse inimesi kirjutama kõigile mõistetavas keeles, lähtudes just keskmisest keeletarbijast. Samas rõhutatakse, et see ei ole robustselt lihtsustatud keel, vaid läbimõeldud ja sisule keskenduv väljendusviis (Hallik, 2011). Näiteks koostab Soome Rahvusringhääling YLE paralleelselt oma Soome-teemalisi uudiseid ka keeleõppijatele („Yle Uutiset selkosuomeksi“, s.a.).

Kui varem on levinud palve, eelkõige just erivajadustega inimesi silmas pidades, sama siuga erinevalt ülesehitatud tekstide koostamiseks, siis nüüd on hakatud keskenduma raketuste loomisele, mis teostaksid lihtsustamist automaatselt. Peamiseks põhjuseks on ühekordse programmi kirjutamise odavus võrreldes tekstide korduva käsitsi koostamisega (Saggion jt, 2015: 2).

1.1. Teksti lihtsustamise viisid

Teksti on võimalik teha lihtsamaks, muutes mõnda selle osa või kõiki osasid. Seejuures tuleks arvesse võtta, et selle tulemusel võib kaduma minna mõni aspekt, mis on sisu või vormi seisukohalt oluline (Chandrasekar jt, 1996: 1041). Teksti raskusaste sõltub suuresti selle lausete pikkusest ning selles esineva sõnavara tavapärasusest (Mikk, 2015). Niisiis võib üldisteks teksti lihtsustamise viisideks nimetada leksikaalset ehk sõnavaralist ja süntaktilist ehk lausestuse lihtsustamist. Suurema mahuga terviktekstide lugemist hõlbustab aga kokkuvõtte tegemine või märksõnade leidmine.

¹ Eestis on selle näiteks selge keele ja selge sõnumi liikumise koduleht <https://selgesonum.ee/>.

1.1.1. Leksikaalne lihtsustamine

Teksti leksikaalse lihtsustamise korral asendatakse vähem levinud sõnad, näiteks võõrsõnad, sagedamini kasutatavate sõnadega, kasutades selleks peamiselt juba loodud andmebaase: sünonüümi- ja sagedussõnastikke ning leksikaalset andmebaasi Eesti Wordnet² (Peedosk, 2017a). Leksikaalne lihtsustamine toetab neid, kellel on raskusi pikkade sõnade lugemisel või kelle sõnavara on kitsas, näiteks keeleõppijaid ja lugemisraskustega inimesi (Sag-gion jt, 2015: 4). Sõnade asendamisel üritatakse leida kõige sarnasemat, kuid igal lekseemil on oma tähendusnüanss, mille kadumine võib muuta kogu teksti emotsiooni.

(1) Platsi *serva* jäi värske kasekänd, millel sobiks istuda, kui ringi keskel löke *leegitseks*.

(1a) Platsi *äärde* jäi värske kasekänd, millel sobiks istuda, kui ringi keskel löke *põleks*.

Ilukirjanduskorpusest pärit näitelause 1 tuvastab Martin Peedoski programm³ muutmisväärses sõnad *serva* ja *leegitseks*. Sõnale *serva* pakutakse asenduseks *äärde* ja *kanti* ning *leegitsema* asendajaks *põleks*, *lõõmaks*, *õhetaks*, *säraks*. Käesoleva töö autori meelest on tulemuseks saadud lause 1a üsna sarnane oodatud lahendusele. Täpsemalt uurides ei selgu algsest lausest, millist põlemise staadiumit on sõna *leegitsema* all silmas peetud, mistõttu sobib antud juhul kõige levinum variant *põlema* väga hästi.

1.1.2. Süntakiline lihtsustamine

Süntaksi lihtsustamise korral muudetakse tekstis lausestruktuuri. Selleks kasutatakse reegleid, mille sõnastamisel on sageli lähtutud süntaksipuust (Shardlow, 2014: 62; Chandrasekar jt, 1996). Keele kasutajale on tavaliselt peamiseks probleemiks liiga pikad laused, mis koosnevad mitmest osalausest, mis omakorda võivad olla mitmel eri tasandil. Lühemates lauses on aga korraga edastatud vähem informatsiooni, mis annab võimaluse keskenduda ühele mõttele korraga (Siddharthan, 2006: 77). Näiteks töö autori poolt moodustatud komplekslause ehk mitut sündmust väljendav lause 2 on võimalik jagada kolmeks eraldi lauseks: 2a, 2b ja 2c.

(2) Kass, kes kõndis mööda teed, nägi hiirt ja hüppas tema poole.

² Andmebaas Eesti Wordnet on leitav aadressilt <https://www.cl.ut.ee/ressursid/teksaurus/?lang=et>.

³ Martin Peedoski loodud sõnavaralihtsustamise veebirakendus on leitav aadressilt <http://prog.keeleressur-sid.ee:4567/>.

(2a) Kass kõndis mööda teed.

(2b) Kass nägi hiirt.

(2c) Kass hüppas hiire poole.

(3) Lõputöö on kirjutatud Ruthi poolt.

(3a) Ruth kirjutas lõputöö.

Lausestuse puhul võib tähelepanu pöörata ka tegumoele (Shardlow, 2014: 62). Mõningatel juhtudel on võimalik muuta lauses passiiv aktiiviks, sest see osutab elusale või elutule tegijale (Erelt, Erelt ja Ross, 2007: 445–457). Näiteks passiivis olevast *poolt*-tarindiga lausest 3 on võimalik moodustada aktiivi lause 3a.

1.1.3. Tekstist põhilise ekstraheerimine

Teksti lihtsustamine kokkuvõtmise meetodil aitab leida peamisi mõtteid lausete või fraaside näol, seejuures sõltub tulemus algteksti pikkusest (Chandrasekar jt, 1996: 1041–1042). Veel on sellise lihtsustamise omaduseks, et tulemuseks võivad olla täiesti uued, kuid sama sisu edasiandvad laused (Müürisep, 2006: 2). Olulise ekstraheerimine aitab lahendada probleemi, et sageli ei mõisteta teksti põhisisu, vaid jäädakse kinni kõrvalistesse aspektidesse. Meetodist on kõige rohkem kasu autismispektri häire(te)ga, lugemis- või keskendumisraskustega inimestele, neid ebavajalikuga mitte koormates (Saggion jt, 2015: 4).

1.2. Teksti automaatne lihtsustamine teiste keelte näitel

Maailmas on loodud erinevaid teksti lihtsustamise programme. Guugeldades inglise keelele tehtud teksti lihtsustavaid programme, leidis töö autor mitmeid erinevaid (ennekõike sõnavarale keskenduvaid) rakendusi, kuid süntaksi kohta vaid ühe võimaluse (vt ptk 1.2.2). Selle üheks põhjuseks võib olla, et teksti sõnavara on lihtsam automaatselt modifitseerida, kuna sõnavarale keskenduvate rakenduste loomist abistavaid sõnastikke on palju. Samal ajal aga süntaksi lihtsustamist takistavad mitmed tegurid: algteksti õigekirja nõue või väga paindlik keelesüsteem, mille tõttu on keerukas automaatselt osalauseid eraldada.

Aastal 2014 tegi Matthew Shardlow vahekokkuvõtte teksti lihtsustamise suundadest erinevates keeltes ning sellest võib näha suuremat teadustasemel huvi just süntaksi vastu. Ainult leksikaalset lihtsustamist oli uuritud 3 erineva keele (jaapani, hispaania ja rootsi keel) ja süntaktilist lihtsustamist 6 keele puhul (taani, prantsuse, vietnami, baski, itaalia ja korea

keel). Brasiilia portugali keele ja itaalia keele puhul oli korruga lihtsustatud mõlemat aspekti (Shardlow, 2014: 66).

1.2.1. Uuringud süntaksi lihtsustamiseks

Uurimusi süntaksi lihtsustamisest on tehtud erinevatel eesmärkidel ja erinevatele sihtrühmadele. Kui üldiselt on mõeldud võõrkeelena kõnelejatele ja lugemisraskusega inimesetele (Saggion jt, 2015), siis teiste hulgas mainitakse veel viipekeele kasutajaid, kes eelistavad lugeda-viibelda lihtlauseid (Aluísio ja Gasperin, 2010). Vaegkuuljatele on suunatud näiteks korea keelele tehtud rakendus, kus lisaks lihtsustamisele osalused ka visualiseeritakse (Chung, Min, Kim ja Park, 2013). Täpsemalt on oma sihtrühma määratlenud ka itaalia keelele mõeldud rakenduse ERNESTA loojad, kes üritasid lihtsustada 7–11-aastastele lastele mõeldud tekste (Barlacchi ja Tonelli, 2013). Mitmel juhul on lihtsustamise vajadus tekkinud masintöötuse tarbeks, näiteks baski ja vietnami keele puhul leiti, et masintõlge saab palju paremini hakkama just lihtlauseetega (Aranzabe, Diaz de Ilarraza ja Gonzalez-Dios, 2012; Hung, Minh ja Shimazu, 2012). Hoopis teises suunas on aga töötatud korruga taani ja inglise keele lihtsustamisel, kus eesmärgiks oli lauseid lühendada selleks, et subtiitrid ekraanile ära mahuksid (Daelemans, Höthker ja Sang, 2004).

Vastavalt sihtrühmale ja eesmärgile erinevad kohati ka lihtsustamiseks võetavad aspektid. Järgnevalt võrreldakse tabelisse 1 kogutud andmete põhjal Brasiilia portugali (toetudes Aluísio ja Gasperin, 2010), itaalia (toetudes Barlacchi ja Tonelli, 2013), baski (toetudes Aranzabe jt, 2013), vietnami (toetudes Hung jt, 2012) ja korea (toetudes Chung jt, 2013) keele kohta välja toodud lihtsustamisaspekte ja -algoritme, mis on kõik reeglipõhised lähenemised. Protsess algab alati osalauseteks tükeldamisest ja süntaksipuu loomisest, kuid mõningatel juhtudel eraldatakse ja märgitakse ka üht või teist tüüpi fraasid ja täiendid.

Eestäienditega tegeletakse eraldi portugali ja korea keele puhul, kus tehakse lisalause täiendist ja selle viitealusust. Baski keele aglutineerivusest lähtuvalt kaotatakse täiendid sõnade küljest hoopis ära. Fraaside jaoks on tegevuste kirjelduste täpsusastmed erinevad. Näiteks portugali keeles liigutatakse liiga pikka määrusfraasi ja vietnami keeles eraldatakse nimisõnafraas eraldi lauseks. Korea keele puhul on välja toodud, et nimisõnafraasist moodustuv lihtlause liigutatakse lause ette, määrusfraasid ja eestäiendid aga lause järele. Kui aga eestäiendina esinev osalause algab mäarsõnaga, siis tuleb sellest ja nimisõnafraasist moodustada eraldi lihtlause lühendatud lause ette, niisamuti nagu ka kaudset kõneviisi sisaldav

Tabel 1. Erinevatele keeltele loodud reeglipõhiste lihtsustamissüsteemide ülevaade

(lihtsustamisliikide lühendid: Sy = süntaks, Le = leksika).

Keel(liik)	Nimi, amlikas	Sihtrühm/vajadus	Turimissalused	Taktika
Brasilia portugali (Sy; Le)	PorSimples (Ahuisio ja Gasperin, 2010)	<ul style="list-style-type: none"> • Veebis dokumendid • Lapsed, eri liiki tekstid • Täiskasvanud • Kuulmishäiretega (viipek) • Kaugõppes 	<ul style="list-style-type: none"> • Eeständ • Sidesõnaga algavad kõralaused • Rindlausend • Põimlausend • Passiiv • Mitte SVO järjega • Pikad määrusfraasid 	<ul style="list-style-type: none"> • Tükeldab lause • Mundab diskursuse lihtsamaks • Passiiv aktiiviks • Mundab osalause järgkorda • Sõnajärg SVO-ks • Liigutab pikka määrusfraasi
Itaalia (Sy)	ERNESTA (Barlacchi ja Tonelli, 2013)	<ul style="list-style-type: none"> • Lapsed 7–11a (madal lugemisoskus) • Muinasjutud 		<ul style="list-style-type: none"> • Leiab anafoorid (tasemed) • Lihtsustab lause-lausel • Jätab ainult faktilise info alles <ul style="list-style-type: none"> • olevikku
Baski (Sy)	(Aranzabe, Iarraza ja Gonzalez-Dios, 2012; 2013)	<ul style="list-style-type: none"> • Masintõlge • Õppimine, välismaalased jne <p>(Keel on aglutineeriv ehk sõnu moodustatakse liidete lisamisel sõnatüvele; sõnajärg on vaba)</p>	<ul style="list-style-type: none"> • Relatiivlausend • Ajamäärusega osalause (ingl <i>adverbial temporal clause</i>) 	<ul style="list-style-type: none"> • Määrab grammatikaga tükelduskohad (osalauseid) • Algoritmiga pundeks • Tükeldab • Rekonstruktsioon: eemaldab rüüendid (aglutineeriv keel -> osa sõnast); lisab ad-verb / sõnastab tümber • Paigutab tümber: lauses sõnad / terved laused • Korrasab grammatika ja õigekirja (suurtäht, punkt) • Teeb sulgudes olevaga eraldi lause
Vietnami (Sy)	(Hung, Minh ja Shimazu, 2012)	<ul style="list-style-type: none"> • Fraasipõhine masintõlge 		<ul style="list-style-type: none"> • Tükeldab osalauseteks • Tükeldab nimisõnafrasi • Järgestab tümber
Korea (Sy)	(Chung, Min, Kim ja Park, 2013)	<ul style="list-style-type: none"> • Vaegkuuljad 	<ul style="list-style-type: none"> • Nimisõnafrasid [ettepoole] • Kaadne kõneviis [ettepoole] • Määrusfrasid [tahapoole] • (Ees)täiendid (ingl <i>prenominal clause</i>): määrusõnaga [ettepoole, võta kaasa nimisõnafrasi]; eeständ [tahapoole] 	<ul style="list-style-type: none"> • Tuvastab kõrvallaused (+tüüp) [vaatusalused fraasid] • Koostab süntaksipuu • Paigutab tümber osalauseid (süntaktiliste ja semantiliste omaduste järgi) • Visuaalseerib osalauseid/(te suhted)

osalause. Portugali keele puhul jälgitakse ka passiivis olevaid lauseid, mis muudetakse aktiiviks. Keelele omast sõnajärge ning lausete loogilist järjestust kontrollitakse portugali, baski, vietnami ja korea keele puhul. Baski keele puhul korrastatakse veel tulemuse grammatikat ja õigekirja, mainitud on ka lause algustähe suureks muutmine ja lause lõppu punkti panek. Teistest keeltest erinevalt toimitakse itaalia keele puhul, kus lihtsustamine viitab veel ka sellele, et alles jäetakse vaid faktilist infot edastavad laused ning kõik need muudetakse olevikku.

1.2.2. Süntaksi lihtsustamise rakendus YATS demo, Simplext

Programm Simplext⁴ on algselt loodud hispaania keelele, suunatuna just erivajadustega inimestele, kuid lõpptulemusena laiendati uurimistööd süvendatult teksti lihtsustamise võimalustele ning loodi algversioon ka inglise keelele („Simplext – Text simplifier“, 2016). Mõlemale keelele mõeldud programmid lihtsustavad sisendi lauseehitust ja sõnavara, jättes tegemata tekstist kokkuvõtte. Sõnavara lihtsustamiseks kasutatakse sõnadevaheliste kauguste leidmist, mis näitab, kui sarnased on sõnad, ning lausestruktuuri muutmisel võetakse eeskujuks reeglipõhine lihtsustamine ja lausete genereerimine (Ferrés, Marimon, Saggion ja AbuRa’ed, 2016). Kuna töö autor hispaania keelt ei valda, analüüsitakse siin inglise keelele tehtud versiooni nimega YATS demo⁵.

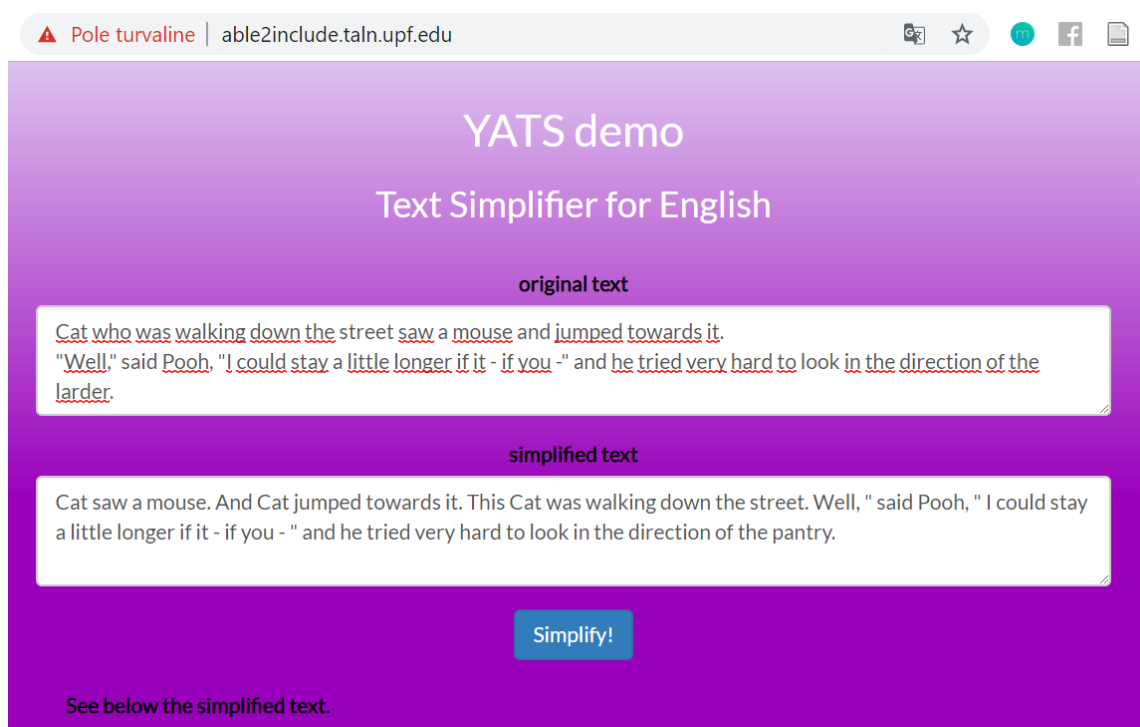
Rakenduse loojad (Ferrés jt, 2016) on programmi tutvustuses välja toonud, et inglise keelele mõeldud versioonis analüüsitakse tekstis esinevaid lausete ja sõnade suhteid, sõnade ees-täiendeid, lauses esinevat passiivi, määrsõnaga osalauseid ning kõiki osalause suhteid. Peatükis 1.2.1 kirjeldatuga sarnaselt alustatakse lausete tükeldamisest, järgnevalt tõstetakse ümber fraase, võetakse arvesse viitesuhteid, võimalusel lihtsustatakse verbi vormi ning lisatakse lause lõppu punkt ja alustatakse lauset suure algustähega. Joonisel 1 on näha veebilehe ülesehitus: teksti saab sisestada selleks ettenähtud esimesse kasti ning tulemus ilmub „Simplify!“ („Lihtsusta!“) nuppu vajutades alumisse kasti.

Rakenduse töö analüüsimiseks tõlgiti peatükis 1.1.2 kasutatud näitelause 2 inglise keelde, sisendlauseks on *Cat who was walking down the street saw a mouse and jumped towards it.*

⁴ Veebirakendus Simplext hispaaniakeelse teksti lihtsustamiseks on leitav aadressilt <http://simplext.taln.upf.edu/>.

⁵ Veebirakendus YATS demo ingliskeelse teksti lihtsustamiseks on leitav aadressilt <http://able2include.taln.upf.edu/>.

Vastuseks kuvati kolmest lausest koosnev väljund: *Cat saw a mouse. And Cat jumped towards it. This Cat was walking down the street.* Sellega võib üsna rahule jääda, ehkki oleks olnud loogilisem alustada hetkel viimaseks jäänud lausega. Samas, kolmanda lause algusesse lisatud sõna *This* ilmumist ei osanud töö autor oodata. Lisaks katsetati programmi komatundlikkust. Kui kaks koma lisati sisendlausesse samasse kohta, nagu need olid eesti-keelses lauses (2), siis oli programmi vastuseks sisendiga samasugune tulem. Juhul, kui koma jätta vaid sõnade *Cat* ja *who* vahele, siis on tulemus kahelauseline: *Cat, who was walking down the street saw a mouse. And the street jumped towards it.* Toetudes asjaolule, et kassi asemel pandi hüppama tänav, võib lugeda tõestatuks, et rakendus on tundlik õigekirja suhtes. Veel võib mainida, et kui sõna *Cat* ette panna artikkel *A* või *The*, on lihtsustamise tulemuseks, sõltumata komade panekust, alati joonisel 1 nähtav kolmelausealine vastus.



Joonis 1. Veebirakenduse YATS demo (2016) kasutusnäide.

Teine joonisel 1 sisendi kastis olev lause *"Well," said Pooh, "I could stay a little longer if it - if you -" and he tried very hard to look in the direction of the larder* on pärit Alan Alexander Milne raamatust „Winnie-the-Pooh“ (Milne, 1981: 23). Selle abil on näha, et otsekõnet sisaldava lausega rakendus väga hästi hakkama ei saa. Näiteks kaotatakse ära alustavad jutu-märgid ja kogu lauseehitus jääb samasuguseks. Küll aga asendatakse sõna *larder* sõnaga *pantry*, mida võib pidada põhjendatuks, toetudes nende kahe sõna leiduvusele Keeleveebis asuvas neljas inglise-eesti sõnaraamatus.

2. Eesti keele automaatseks analüüsiks loodud tehnilised vahendid

Eesti keele jaoks on koostatud ja on koostamisel mitmeid tehnilisi vahendeid, mille abil on võimalik automaatselt tekste analüüsida. Paralleelselt luuakse kasutajaliideseid erinevate võimaluste eraldiseisvaks kasutuseks veebis ning arendatakse Pythoni teeki mugavaks programmeerimiseks. Arvutilingvistidele ja keeletehnoloogidele on kõik loodud rakendused ühendatud teeki EstNLTK, mida arendatakse Tartu Ülikoolis Sven Lauri juhtimisel (Laur, i.a.).

2.1. EstNLTK

Projekti EstNLTK (Laur, i.a.) käigus on loodud mitmeid vahendeid eesti keele kui loomuliku keele töötamiseks. Teiste hulgas on tähtsaim EstNLTK teek, mis on mõeldud programmeerimiskeelele Python ning sisaldab peamisi loomuliku keele töötamiseks vajalikke funktsioone. Selle abil on võimalik tuvastada sõna- ja (osa)lausepiire, leida ja moodustada sõnavorme ning määrata sõnaliike. Lisaks saab teostada pindsüntaktilist ja sõltuvussüntaktilist analüüsi. Viimastes versioonides on võimalik automaatselt leida ka nimeolemeid ja ajaväljendeid ning loodud on mugav ühendus Eesti Wordneti ja Vikipeediaga (*ibid.*).

Teised projekti („EstNLTK dokumentatsioon“, i.a.) osana arendatud tööriistad aitavad leida eestikeelsest tekstist konkreetset infot või teatud andmeid visualiseerida. Veebitööriist NER-tagger on mõeldud nimeolemite automaatseks märgendamiseks ja teksti annoteerimiseks; veebitööriista Text-viewer abil saab tekste etteantud mustrite järgi märgendada või loodud eraldusmustreid testida; Textclassifier on mõeldud andmete kategoriseerimiseks. Tööriistade Cluster-labelling-plot ja Pattern-examiner abil saab tekstiosasid klasterdada ja võrrelda ning hinnata tekstide sarnasust. Tekstis tihedamini esinevaid sõna- ja täheühendeid võimaldab leida ja rühmitada Episode-miner ning kasutajale pakub lauses lünga täitmiseks valikuid veebipõhine tööriist Gap-tagger. Kahe csv-sagedusloendi visuaalse võrdluse kuvab käsuraatööriist Volcanoplot (*ibid.*).

Siinses töös on võtmetähtsusega teksti osalused ning seetõttu on vajalik nende automaatne eraldamine. Tartu Ülikooli teadlaste Heiki-Jaan Kaalepi ja Kadri Muischneki (2012) loodud osalauseste eraldaja suudab eristada algsest lausest erinevat liiki osalauseid. Lisaks eraldab see kiile, mis osalausestajat tutvustavas artiklis (*ibid.*: 58) on defineeritud järgmiselt: „Kiil on siinses käsitluses tervikuna teise süntaktilise üksuse sees paiknev süntaktiline üksus, s.t.

sama süntaktiline struktuur algab enne kiilu ja jätkub pärast kiilu lõppu.“ Samas jätab programm aga määramata osalause te tasemed ning nende süntaktilised omadused. Programmi tööpõhimõte on liikuda lihtsamast raskemate juhtumite poole (Kaalep ja Muischnek, 2012). Antud moodul on teegi EstNLTK osa ning eraldiseisvana (näiteks veebirakenduses) seda kasutada ei saa.

2.2. Eesti keele süntaksianalüsaator

Lausete automaattöötlemises on oluline ka süntaksianalüüs. Selle jaoks töötasid Kaili Müürisep ja Tiina Puolakainen riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)“ raames välja süntaksianalüsaatori („Eesti keele süntaksianalüsaator“, i.a.). Kaili Müürisep on selle temaga põhjalikumalt tegelenud ka oma magistri- ja doktoritöös (1996; 2000).

Programmi tööpõhimõte on kõigepealt leida sõnadele eraldi nende kõikvõimalikud funktsioonid ning hiljem võtta aluseks kontekst ja eemaldada nimekirjast ebasobilik (Müürisep, 2004: 207). Analüsaatorit on võimalik kasutada lisaks kirjakeelele ka suulise kõne transkriptsioonidel (Müürisep, Nigol ja Uiibo, 2006). Peale EstNLTK teegi osana programmeerimiskeskonnades saab seda kasutada veebiversioonis⁶, viimases näeb paralleelselt märgendusega tulemust ka joonisel.

2.3. Eesti keele morfoloogiline analüsaator, ühestaja ja süntesaator

Morfoloogiaanalüsaatorit ja spellerit on arendatud juba aastast 1991, seda teemat on peamiselt käsitlenud Heiki-Jaan Kaalep (1998). Arvutilingvistikale spetsialiseerunud firma FiloSoft OÜ on oma kodulehel kättesaadavaks teinud ka teisi internetis kasutatavaid programme („FiloSoft“, i.a.). Sealsamas on olemas nii morfoanalüsaator kui ka morfosüntesaator. Need kaks ja morfoloogilise analüüsi ühestaja on hiljem loodud EstNLTK teegi osad.

Morfoloogilise analüsaatori⁷ peamiseks eesmärgiks on leida teksti osadele analüüs, kus määratakse selle algvorm, sõnaliik ja kasutatav käänne või pööre ning eraldatakse tüvi ja

⁶ Süntaksianalüsaator on leitav aadressilt <https://korpused.keeleressursid.ee/syntaks/index.php>.

⁷ Eesti keele morfoanalüsaator on leitav FiloSoft OÜ veebilehel http://www.filosoft.ee/html_morf_et/.

käände- või pöördelõpp. Liitsõnade ja tuletiste puhul analüüsitakse iga nende osa veel ka eraldi (Kaalep, 1998: 24–26). Morfoanalüsaator võib leida ühele tekstis esinevale sõnale mitu erinevat analüüsi, millest õige valiku langetamiseks on reeglina abiks kontekst. Selle lahendamiseks, milline märgend sõnale jätta, on Heiki-Jaan Kaalep ja Tarmo Vaino (1998) loonud ühestaja, mida rakendatakse just mitmese analüüsiga sõnadele. Filosoofi loodud ühestaja kasutab sobivaima analüüsi leidmiseks reegleid tõenäosuslikust mudelist nimetusega Markovi varjatud mudel (*ibid.*).

Morfoloogiline süntesaator⁸ teeb ühestajale vastupidist tööd, programmi sisendiks on sõna algvorm, mille muutevormi soovitakse. Käändsõna puhul tuleb ära märkida soovitatav kääne ning arv, pöörsõna puhul isik, aeg ja kõneviis või käändelise vormi ootuspärasus. Väljundiks on vastavalt valikutele genereeritud grammatiline vorm.

⁸ Eesti keele süntesaator on leitav Filosoofi OÜ veebilehel http://www.filosoofi.ee/gene_et/.

3. Eestikeelse teksti lauseehitust lihtsustav rakendus

Bakalaureusetöö käigus loodi rakendus, mille abil on reeglipõhiselt võimalik lihtsustada eestikeelse teksti lausestruktuuri. Rakendus on kättesaadav Eesti Keeleressursside Keskuse veebiserveris⁹ ja programmi kood on avalik keskkonnas GitHub¹⁰.

Koodide käivitamiseks väljaspool serverit on vajalikud Python, versioon 3.5, ning EstNLTK teek, versioon 1.4.1. Programmeerimiseks valiti Python, sest just selles keeles on kõige mugavam kasutada EstNLTK teeki, milles on olemas võimalused eestikeelsete lausete mitmekülgseks automaatseks analüüsiks (vt ptk 2). Teegi versioonidest kasutati vanemat, sest uuem on arendamisjärgus.

3.1. Rakenduse prototüüp

Prototüüp süntaksi lihtsustamise programmile valmis aines „Eesti keele töötlus Pythonis“. Selle eesmärgiks oli kontrollida, kas osalausestaja abil saab muuta liitlauset lihtlauseteks.

Uurimiseks võeti juhuslikkuse alusel korpusest ilukirjandustekst ning ajakirjanduskorpusest üks Postimehe ja üks Päevalehe tekst. XML-kujul failidest võeti tekst lõikude kaupa ehk märgendite $\langle p \rangle$ ja $\langle /p \rangle$ vahelt. Iga lõik tehti EstNLTK abil *Text*-objektiks, edasi leiti selle osalused. Osalused ja lõigud koguti listidesse ning iga algse faili kohta tehti uus fail, mille sisuks olid antud listid. Uued failid lisati kausta „korpusfailid“. Selline eeltöötlus tagas hili-semaks tegutsemiseks kiirema analüüsiaja, sest osalauseid ei pidanud enam iga kord uuesti genereerima. Samadele tekstidele toetuti hiljem ka bakalaureusetöö käigus, algsed tekstid on nähtavad GitHubis¹⁰ ning siin töös kasutatud näitelauseid on pärit just sellest korpusest.

Prototüübis tükeldati otsekõnet mitte sisaldav liitlause ja kohendati iga osa nii, et järgnevalt kirjeldatud tingimuste täitmisel pole need enam väljajätelised. Alati eemaldati algusest ja lõpust ebasobilikud kirjavahemärgid ja sidesõnad, vajadusel lisati osalause lõppu lauset lõpetav märk ning muudeti lause algustäht suureks. Verbi või ebaselge pronoomeniga lausete puhul üritati leida verbi või esialgse asesõna abil sobilik asesõna lause algusesse. Järgnevalt kasutati programmi kohandatud korpustekstidel ning saadud tulemused liigitati võimaliku

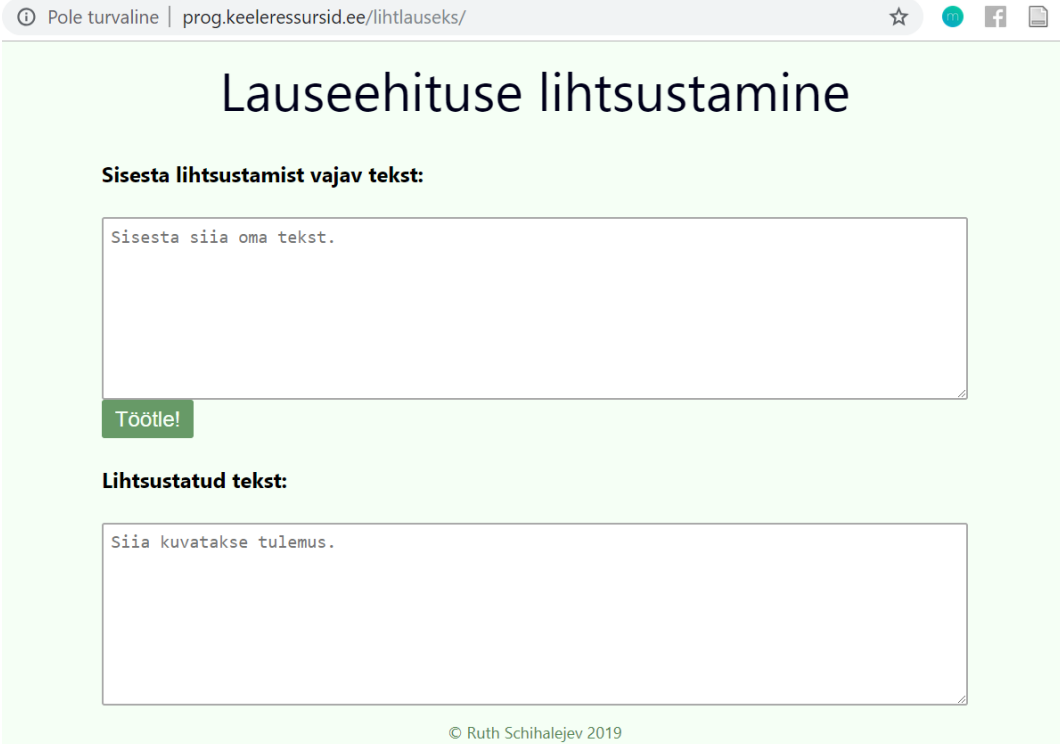
⁹ Loodud rakendus on Eesti Keeleressursside Keskuse veebiserveris <http://prog.keeleressursid.ee/lihtlauseks/>.

¹⁰ Loodud rakenduse kood on GitHubis <https://github.com/ruthsch/lauseehituse-lihtsustamine>.

parandusvariandi järgi gruppidesse. Selle bakalaureusetöö käigus genereeriti induktiivselt täpsemad reeglid, pandi nad kirja ning analüüsiti, kas need täidavad oma eesmärgi.

3.2. Veebirakenduse ülevaade

Veebirakenduse kujundamisel võeti eeskujuks eelkirjeldatud YATS demo (2016) ja Martin Peedoski bakalaureusetöö (2017a) raames loodud sõnavaralihtsustamise programm (Peedosk, 2017b). Toetudes töö autori kasutuskogemusele, on nende etalonina kasutamise põhjuseks selge disain ja kasutajasõbralikkus. Nagu joonisel 2 on näha, paikneb lehel kaks tekstikasti. Ülemisse välja saab sisestada oma teksti, mille võib sinna kopeerida või ise trükkida, ning vajutades nupule „Töötle!“ kuvatakse tulemus allpool olevasse kasti. Väljundi koostamise ja ilmumise aeg oleneb sisendi pikkusest.



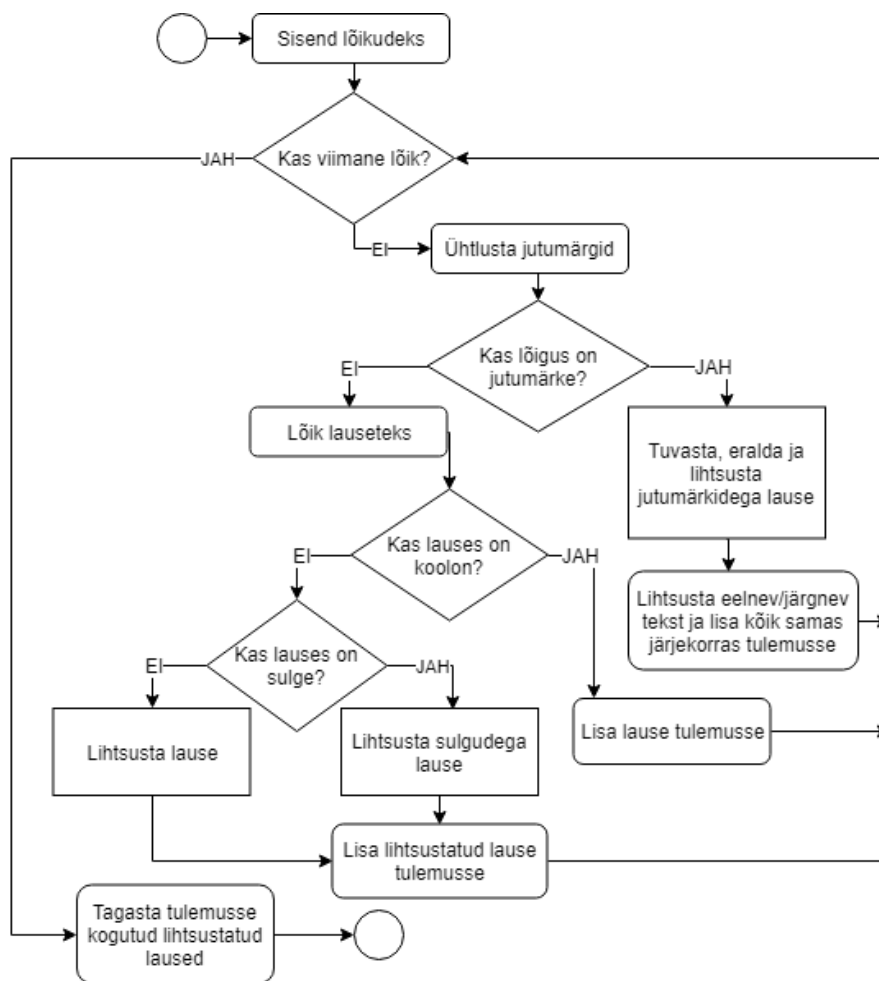
The screenshot shows a web browser window with the address bar displaying 'prog.keeleressursid.ee/lihtlauseks/'. The main content area has a light green background and is titled 'Lauseehituse lihtsustamine'. Below the title, there is a section labeled 'Sisesta lihtsustamist vajav tekst:' followed by a large text input field containing the placeholder text 'Sisesta siia oma tekst.'. Below the input field is a green button with the text 'Töötle!'. Underneath the button is another section labeled 'Lihtsustatud tekst:' followed by a large text output field containing the placeholder text 'Siia kuvatakse tulemus.'. At the bottom of the page, there is a small copyright notice: '© Ruth Schihalejev 2019'.

Joonis 2. Veebirakenduse kujundus.

Veebileht on struktureeritud HTML-iga (ingl *Hypertext Markup Language*) ja kujundatud CSS-iga (ingl *Cascading Style Sheets*). Infovahetus HTML-i ja lihtsustamisprogrammi vahel toimub PHP (ingl *Hypertext Preprocessor*) abil.

3.3. Lihtsustamisprogrammi algoritm

Lihtsustamise programm koosneb etteantud reeglitest, mis on koostatud analüüsi tulemusena. Joonisel 3 on näha, milline on sisendteksti analüüsi järjekord, et vältida teatud tüüpi kirjavihemärkide vahel oleva info nihkumist ebasobivasse kohta. Teksti analüüsitakse lõikude kaupa, tükeldatuna reavahetusele viitava kombinatsiooni $\backslash n$ kohalt, sest nii on võimalik säilitada algse tekstiga sama üldstruktuuri. Iga lõigu lausete lihtsustatud variandid kogutakse kokku ühte *string*-tüüpi muutujasse, mis hiljem tagastatakse.



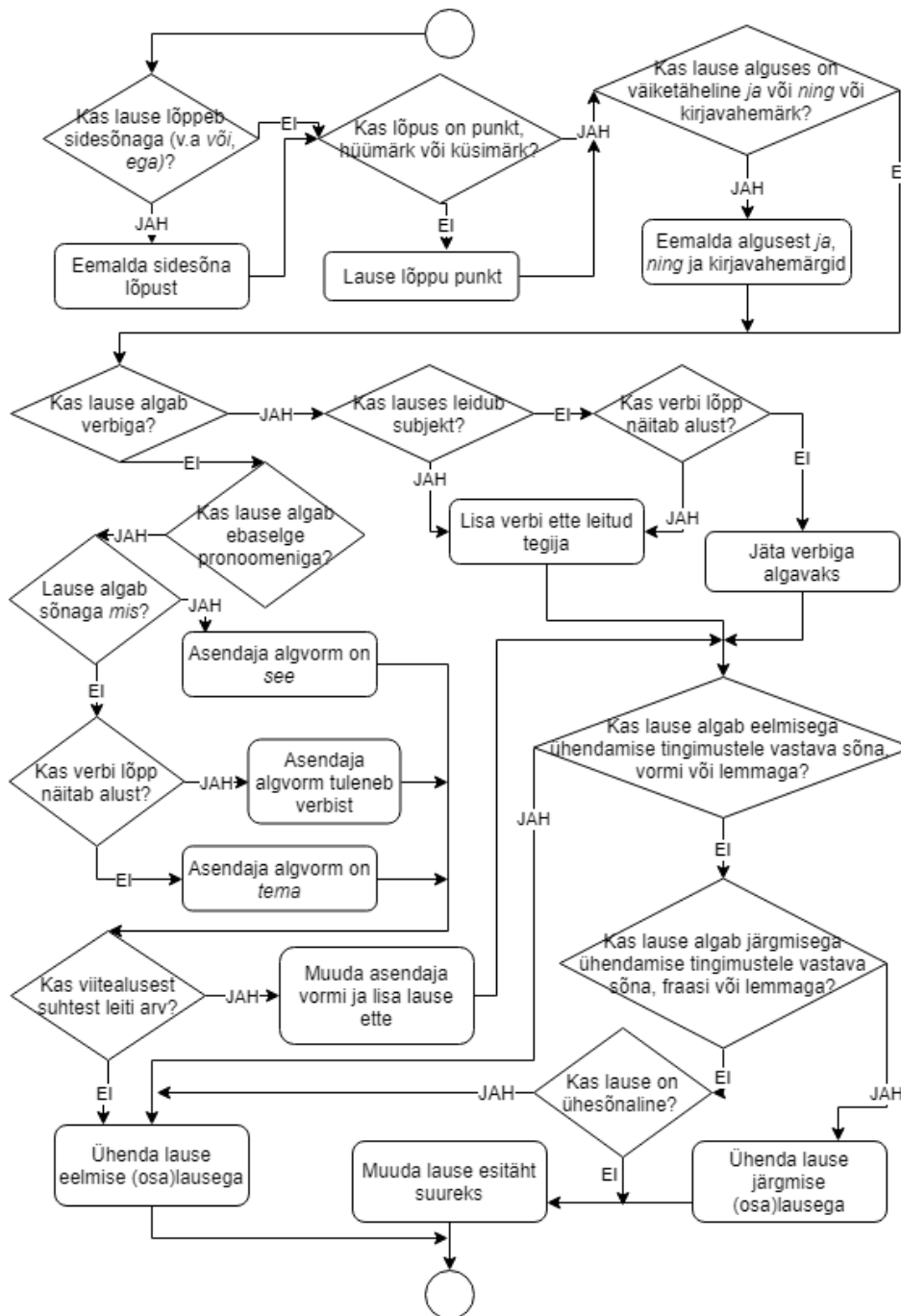
Joonis 3. Lihtsustamisprogrammi peaalgoritm.

Erinevates allikates võib kohata isesuguseid jutumärke, sest nende kasutusnõuded on aegade jooksul palju kõikunud (Raadik, 2014). Automaattöö lihtsustamiseks ühtlustatakse kõik tüübid sirgjutumärkideks (" "). Esimesena kontrollitaksegi jutumärkide olemasolu lõigus, sest jutumärkide sees võib samuti olla mitu lauset, mida lauseteks ja osalauseteks jagamisel ei eristata pealausest. Jutumärkidega lõigu analüüsist saab täpsemalt lugeda peatükis 3.3.2.

Kui lõigus jutumärke ei leidunud, tehakse sellest EstNLTK abil *Text*-objekt ja vaadeldakse iga lauset eraldi. Koolonit sisaldav lause tuleb jätta samale kujule, sest vastasel korral võib juhtuda, et mõni loetelu osa arvestatakse eraldi osalauseks ning mõtte terviklikkus saab tugevalt kannatada. Järgnevalt kontrollitakse sulgude leidumist lauses, sest nendes olev info on sageli täpsustus teatud sõnale ning seega tuleb see hoida täpselt samas kohas. Selleks eraldatakse sulgudes osa(d) lausest ning lause lihtsustatakse. Sulgudes info lisatakse tulemusteksti tagasi, lähtudes sellele eelnenud või järgnenud sümbolitest, mis tuvastatakse regulaaravaldise abil. Eelistatum on variant, et tuginetakse sulgudele eelnenud osale, sest reeglina on sulgudes info pigem osalause keskel või lõpus. Kui aga see ei õnnestu, lisatakse sulgudes olev info sellele järgnenud sümbolite järgi. Viimane realiseerub juhul, kui (osa)lause algab sulgudega. Kui nüüdseks ei ole lauset veel mingi kindla sümboli järgi lihtsustatud, siis lihtsustatakse seda järgnevas alapeatükis (ptk 3.3.1) kirjeldatava algoritmi järgi.

3.3.1. Lause lihtsustamine

Programmi keskne meetod on lausete lihtsustamine ning neid käsitletakse ükshaaval, et osalauseste eraldamisel ei läheks algsed lausepiirid kaduma. Peatükis 2.1 kirjeldatud osalausestaja abil saadud lauseosad analüüsitakse ühe kaupa, toetudes vajadusel ka algele terviklausele. Eesmärk on igast osalausest teha omaette täielik lause, kuid mõnel juhul ei ole see teksti sidususe ja selguse huvides võimalik. Ühe osalause analüüsimise algoritm on toodud joonisel 4.



Joonis 4. Osalause lihtsustamise algoritm.

Kõigepealt uuritakse lause lõpus ja alguses olevaid kirjavahemärke ning sidosõnu. Kirjavahemärgid võib lause algusest kõik ära võtta, lõpus tuleb aga vajadusel asendada sobiliku variandiga. Lause alguses olevatest sidosõnadest tohib ära võtta vaid väiketähega algavad *ja* ja *ning*, lõpu jäävatest sidosõnadest tuleb ära võtta kõik peale sõnade *või* ja *ega*.

Kui puhastatud lause algab verbiga, siis tuleks leida sellele ka tegija. Esimesena otsitakse alust osalause liikmetest, lähtudes selle osalause ja terviklause süntaksianalüüsist. Täpse-

malt huvitavad siinkohal juhud, kus osalause subjekti rollis olev sõna on pikas lauses objekt. Näiteks ajakirjanduskorpusest pärit lausekatke *sai viga nina* on sobilik viia kujule *Nina sai viga*. Kui objekti rollis oleval sõnal on ka eestäiendeid, viiakse verbi ette terve fraas. Kui lause seest alust ei leitud, proovitakse see määrata verbilõpu järgi. Selleks leetakse verbilõpud, mis viitavad ainult ühele võimalikule pronoomenile. Näiteks verbi *tegime* lõpus olev *-me* kinnitab, et pronoomeniks on *meie*. Kui aga kummalgi moel tegijat ei leitud, jäetakse lause algus samaks.

Kui lause verbiga ei alanud, kontrollitakse juhtu, et see algab pronoomeniga, mis ei ole piisavalt selge. Näiteks ei asendata sõnu algvormiga *mina, sina, tema, meie, teie, nemad, see, too, kõik, ise, mõni, sama, säärane, mõlema, üks, iga, keegi, selline, oma, kumb, mitu, miski, üksteise, missugune*. Kui alguses on sõna algvormiga *mis*, võetakse asendajaks *see*. Muul juhul üritatakse tegija määrata eelnevalt kirjeldatud moel verbi järgi, kuid selle nurjumisel võetakse asendajaks *tema*. Järgnevalt on vaja panna asendaja lause konteksti sobivasse vormi, mille määramisel lähtutakse algse pronoomeni käändest ning võimalusel viitealuse sõna arvukategooriast. Viitealuste suhete automaatse leidmisega tegeleb hetkel magistrant Linda Freienthal, kelle bakalaureusetööst (2018) võeti mõned väljapakutud reeglid. Teiste hulgast valiti kasutamiseks järgnevad asendussõnade *kes* ja *mis* kohta käivad seaduspärad: viitealuseid tuleb otsida samast lausest, eelistada võib substantiiv-kandidaate ning eelistada võib kandidaate, mis on asendussõnast kaks kuni kuus või seitse sõnet eespool (*ibid.*: 44–45). Neid kohandati selle töö jaoks nii, et alati kontrollitakse pikas lauses asendussõna ees olevale komale eelneva sõna liigi vastavust substantiivile, pärisnimele või pronoomenile ning võetakse selle sõna arvukategooria. Kui sobilik viitealune sõna leitakse, siis muudetakse asendussõna vorm morfoloogilise süntesaatori abil ja asendatakse algne pronoomen. Vastasel korral ühendatakse see osalause eelmise analüüsitud osalausega.

Korpusanalüüsi tulemusel leiti veel mitmeid juhte, kui osalause ei tohi eraldada talle eelnevast või järgnevast osalausest. Eelmise lausega koos hoidmine sõltub lause alguses olevatest sõnadest. Eelmise osalausega tuleb komaga eraldatult tagasi kokku liita *des-* ja *mata-*lauselühendiga või sõnaga *kuhu* algavad laused. Sama tuleb teha, kui lause algussõna algvormiks on *kui, kas, kuidas, kus, kust, et, kuid, nagu, missugune, ehkki, ent, kuigi, ehk, miks, kuni, mil, vaid, sest, vist, aga, kuna, siis, kuivõrd, alates, mitu*. Järgmise lausega kokku liitmine sõltub nii lause alguses kui ka lõpus olevast sõnast või fraasist. Sõnaga *kui, kuna(s), nagu, et* või fraasiga, mis sisaldab sõna *kui*, algav lause tuleb hoida koos sellele järgneva osalausega. Samuti tuleb toimida, kui lause lõppeb sõnaga *see, seda, või* või *ega*. Kui osalause on

jäänud ühesõnaliseks, tuleb see võimalusel liita *ja*-ga eraldatuna eelmise lausega või eelmise puudumisel viia järgmise osalause algusesse.

Vajadusel tuleb lause algustäht muuta suureks ning lisada väljundi koostamiseks mõeldud muutujasse. Kui kõik osalused on läbi vaadatud, tagastatakse algse pika lause lihtsustatud osalused sõnena.

3.3.2. Jutumärke sisaldava lõigu analüüs

Jutumärke sisaldava lõigu korral on vajalik leida ja eraldada muust tekstist jutumärke sisaldav lause ning määrata, kas jutumärgid viitavad otsekõnele, pealkirjale või tsitaadile. Lausete eraldamisel säilitatakse sisendteksti lausete järjekord.

Üldine eesmärk on viia otsekõnet sisaldavad laused kujule, kus saatelause eelneb otsekõnele, näiteks *X ütles: „...“*. Saatelause ja otsekõne eristamiseks toetuti „Eesti keele käsiraamatus“ (Erelt jt, 2007: 569–570) välja toodud reeglitele, kus on kindlaks määratud nii kirjavahemärkide asukohad kui ka suur- ja väiketähe kasutus. Regulaaravaldiste abil leitakse üles juhud, kus saatelause on otsekõne ees, saatelause on mitme otsekõne vahel või saatelause järgneb otsekõnele. Eraldatud saatelauses tuleb vajadusel ja võimalusel muuta sõnajärge, mis eesti keeles on tavaliselt SVX ehk alus, öeldis ja muud lause osad. Nimetatud osad üritatakse süntaktilise analüüsi abil leida ja järjestada. Kui aga saatelausest alust ei leita, jäetakse saatelause algsele kujule. Korrastatud sõnajärgjega saatelause ja jutumärkidest puhastatud otsekõne lihtsustatakse eraldi ning lõpuks pannakse nad üheks terviklikuks lauseks kokku tagasi.

Kui regulaaravaldiste abil tehakse kindlaks, et jutumärkides on tsitaat või pealkiri, eemaldatakse see lausest ning lihtsustatakse allesjäänud lauset. Seejärel pannakse jutumärkides info tagasi lausesse, lähtudes sellele eelnenud või järgnenud tähemärkidest. Juhul, kui tagasipanek ei õnnestu, tagastatakse algne lause. Viimane võib juhtuda näiteks siis, kui jutumärkides oli kogu lõigu tekst.

3.4. Rakendusele antud tagasiside

Keele kasutamise jaoks on olemas küll mitmed reeglid ja normid, kuid kohati on need väga paindlikud. Kuna keelekasutuse lihtsust ei ole võimalik kindlalt mõõta valemi või programmi abil, vaid see põhineb subjektiivsel hinnangul, siis selle töö praktilise osana valminud veebirakenduse hindamiseks viidi läbi küsitlus.

Nagu võib näha lisas 1, koosneb küsimustik neljast osast. Esimeses osas küsitakse hinnangut programmi tulemusel muutunud lausetele ja teises muutumatuks jäänud lausetele. Kolmandas osas saab osaleja tagasisidestada enda proovitud lauseid ja veebirakenduse kasutust üldiselt ning neljandas kogutakse osaleja taustaandmed.

Küsitlusele vastas 20 inimest, kelle hulgas oli kõige rohkem 19–30-aastaseid (75%), kuni 18-aastaseid oli 2 (10%) ja üle 30-aastaseid oli 3 (15%). Vanust küsiti vastajate vanuselise profiili teadasaamiseks ning kuigi küsitlust jagati ka välismaal elavate inimestega, siis kõik vastanud on määratlenud oma igapäevaseks suhtlemis- ja lugemiskeeleks eesti keele. Järgnevalt analüüsitakse vastuseid iga etteantud teksti kohta eraldi.

3.4.1. Hinnang muudetud lausetele

Küsitluse esimeses osas olid toodud korpusest pärit laused ning nende tulemustekst veebirakendusest. Näitelause valimisel lähtuti mõttest, et iga reeglibloki kohta oleks vähemalt üks näide, kuid kasutajale ei tutvustatud, mida ühe või teise lause puhul täpselt tähele panna.

(4) *Naisel purunes kulm ja sai viga nina, kaasa sõitnud autoomanikust noor mees jäi terveks.*

(4a) *Naisel purunes kulm. Viga sai nina. Kaasa sõitnud autoomanikust noor mees jäi terveks.*

Esimese teksti (näitelause 4 ja tulemustekst 4a) puhul näidati komplekslause käsitlemist ning olukorda, kus alglauses objektina esinev lauseliige tuuakse verbi ette. Kõige üksmeelsemalt leiti, et sisu on algse tekstiga sama, kuid kõige vähem valiti vastusevarianti, et tulemuses rõhutatakse algse lausega sama kohta. Lisaks arvas üks vastajatest, et ükski pakutud väidetest tulemuslause kohta ei kehti. Tema ja mitmed teised pakkusid välja, et tulemuseks võiks olla kaks lauset, mis on saadud algse lauses koma juurest tükeldades. Lisati, et etteantud tulemuse teises lauses võiks täpsustada, et just naise nina oli see, mis viga sai.

(5) *Aga härra Kallion oli alles noor jurist ning ihaldas kiiresti saavutada pealinna üleviimist.*

(5a) *Aga härra Kallion oli alles noor jurist. Tema ihaldas kiiresti saavutada pealinna üleviimist.*

Teise näiteteksti (näitelause 5 ja tulemustekst 5a) puhul on näha, et suurtähega algavad si-desõnad jäetakse alles ja verbi järgi leitakse asesõna *Tema*. Selle juures leidsid kõik peale ühe, et tulemus edastab sama sisu ning kõik laused on eesti keele grammatikale vastavad. Veel leiti, et on aru saada, kellele asesõna viitab ja et lausete järjekord on loogiline. Kõige vähem nõustuti, et lause on algsest lihtsam. Mitmed tõid välja, et *tema* asemel oleks sobilikum kasutada lühemat vormi *ta*. Märgiti ka seda, et algne lause oli ladusam ja sidusam, ning et lühikesi lauseid ei ole mõtet rohkem tükeldada.

(6) Kõiki müügikohti haldab ainult üks omanik, *mis* tähendab, et praeguses ühiskonnas enesestmõistetav vaba konkurents Elva linna kõige lootustandvamal äripinnal ei toimi.

(6a) Kõiki müügikohti haldab ainult üks omanik. *See* tähendab, et praeguses ühiskonnas enesestmõistetav vaba konkurents Elva linna kõige lootustandvamal äripinnal ei toimi.

(6b) Elva linna kõige lootustandvamal äripinna kõiki müügikohti haldab ainult üks omanik. *See* tähendab, et praeguses ühiskonnas enesestmõistetav vaba konkurents siin ei toimi.

Kolmas tekst (näitelause 6 ja tulemustekst 6a) näitas olukorda, kus sõnaga *mis* algav osalause tehakse eraldi lauseks. Selle puhul valiti kõige rohkem vastusevariante, et sisult on algne ja tulemustekst samad, ning et asesõnade viitealused on arusaadavad ja sõnajärg loogiline. Kommentaare oli vähe ja kõik tõid välja erinevaid aspekte. Üks vastajatest leidis, et lauset on muudetuna just kergem lugeda, kuid teine tõdes, et tulemuses teine lause on ikka sama raskesti loetav. Märgiti ka seda, et rõhulise koha muutus toob lause mõtte paremini välja. Lihtsustamise tulemuseks pakuti välja lause 6b, kuid kuna seal on lauseosi palju ümber tõstetud, näib see automaatse lihtsustamise puhul keeruline teostada.

(7) *Kuna need, kes pomme panevad, ei ähvarda, ja need, kes ähvardavad, ei pommita*, siis näib pommikoerte jooksutamine iga idioodi telefonikõne peale mõttetu tegevusena.

(7a) *Kuna need ei ähvarda, kes pomme panevad, ja need ei pommita, kes ähvardavad*, siis näib pommikoerte jooksutamine iga idioodi telefonikõne peale mõttetu tegevusena.

(7b) Pommipanijad ei ähvarda. Ähvardajad ei pommita. Seepärast näib iga idioodi telefonikõne peale pommikoerte jooksutamine mõttetu tegevusena.

Neljas tekst (näitelause 7 ja tulemuslause 7a) demonstreerib olukorda, kus sõnaga *kes* algav osalause liidetakse tagasi kokku eelmisega. Küsitavust tekitab see, et algelt kiiluna esinenud osalause ei paigutu tagasiühendamisel enam samasse kohta. Probleemi olemasolu kajastub ka vastustes, kus kõige vähem valiti variante, et sõna- või lausejärg on loogiline. Kommentaarides nimetati samuti, et tulemuses ei ole kõrvallause asukoht grammatiliselt õige, kuigi üks tõi välja, et praegust tulemuslauset on parem haarata. Lisaks rõhutati, et algne lause on kergemini mõistav ja mugavamini loetav. Lahenduseks pakuti sõna *siis* kaotamist ning kolmeks eraldi lauseks tegemist. Viimase puhul tõi kolmas kommenteerija välja ka tulemuse, mis on toodud näitena 7b.

(8) Olin juba ligi aasta vabakutseline *ning pidin* rohkem ridasid paberile nõiduma kui palgalised *ajakirjanikud, kes võisid* endale vaheldust leida baarides istumise ja komandeeringutes käimisega.

(8a) Olin juba ligi aasta vabakutseline. *Mina pidin* rohkem ridasid paberile nõiduma kui palgalised ajakirjanikud. *Nemad võisid* endale vaheldust leida baarides istumise ja komandeeringutes käimisega.

Viienda tekstiga (näitelause 8 ja tulemustekst 8a) tuleb välja, kuidas lause tükeldamisel leitakse verbi järgi tegijaid. Selle tulemusteksti kohta nõustusid vastajad ühel meelel, et sisu on algsega sama; tulemus on algsest lihtsam; sõna- ja lausejärg on loogiline; asesõnade viitealused on aru saadavad ja ka kõik laused on eesti keele grammatikale vastavad. Arvamused lahknesid selles osas, et kas tekstis rõhutatakse ikka algsega sama kohta. Kuigi oli üksikuid, kes kommenteerisid, et nad eelistaksid asesõna lühemaid vorme, siis üks vastajatest arvas hoopis, et pikkade kasutamisel tekkiv rõhulisus toob lausete mõtted selgemalt välja.

(9) Tarmo Kahera, *kes* lesis kollase kambri diivanil televiisori vastas, ootas mult vastust.

(9a) Tarmo Kahera ootas mult vastust. *Tema* lesis kollase kambri diivanil televiisori vastas.

Kuuendas tekstis (näitelause 9 ja tulemustekst 9a) on näha olukord, kus kiilust tehti antud lause järele uus lause. Vastajad nõustusid, et tulemus on sisu poolest algsega sama, asesõnade viitealused on mõistetavad ning laused on eesti keele grammatikale vastavad. Kõige vähem valiti vastusevariante selle kohta, et algsega oleks rõhutatud sama kohta, ja et lausete järjekord oleks loogiline. Välja pakuti lausete järjekorra vahetamist või teise tulemuslause ära jätmist. Olukorra lahenduseks arvati olevat ka see, kui kasutada *Tema* asemel *Ta*. Veel

jäädi lahendust pakkumata arvamusele, et algne tekst on ladusam, sidusam, loogilisem ja ilusam lugeda.

(10) Wallingatan meenutab kangesti Tallinna vanalinna, *tundus, nagu* asuaksin Rataskaevu ja Pika tänava nurgal.

(10a) Wallingatan meenutab kangesti Tallinna vanalinna *ja tundus, nagu* asuaksin Rataskaevu ja Pika tänava nurgal.

Seitsmenda tekstiga (näitelause 10 ja tulemuslause 10a) näidatakse olukorda, kus üksik te- gusõna liidetakse eelmise osalausega, lisades nende vahele sidesõna *ja*, ning eelmise osalau- sega ühendatakse sõnaga *nagu* algav lausekatke. Kuna muudatus on väike, siis ei märgitud väga palju vastusevarianti, et tulemuslause oleks algsest lihtsam. Samas tõdeti, et raskemaks see lause ka ei muutunud. Lihtsustamiseks pakuti võimalust tulemuslauses *ja* kohalt kaheks lauseks tegemise ning siis teises lauses alguseks hoopis *Mulle tundus, et asun...*

(11) Oli vaja kraavikaldalt metsa keerata - minu arusaamist mööda ei viinud veenire talu, vaid jõe poole.

(11a) Oli vaja kraavikaldalt metsa keerata. Minu arusaamist mööda ei viinud vee- nire talu, vaid jõe poole.

(11b) Oli vaja kraavikaldalt metsa keerata. Minu arusaamist mööda ei viinud vee- nire talu poole. See viis jõe poole.

Kaheksandas tekstis (näitelause 11 ja tulemustekst 11a) on algselt olukord, kus esimesele osalausele järgneb mõttekriipsu abil täpsustus, kuid tulemuses on need eraldatud kaheks lauseks. Kommenteerijaid oli vaid üksikuid, kuid neist kaks vastandusid omavahel, sest üks arvas, et mõttekriipsu vahetamine punkti vastu ei muuda midagi olulist ning teine leidis, et sellise lahenduse puhul läheb kaduma mõte, et teine lausepool oli selgituse rollis. Leiti veel, et vaatamata mittelihtsustumisele ei ole tulemus raskem ning lahenduseks pakuti algse lause kolmeks jagamist (näitetekst 11b).

(12) "Huvi meie kauba vastu tõuseb," kinnitas Eesti majandusesindaja Venemaal Anne Mar- japuu. "Septembris oli tõesti vaikne, kuid nüüd on küsimisi igasuguseid."

(12a) Eesti majandusesindaja Venemaal Anne Marjapuu kinnitas: "Huvi meie kauba vastu tõuseb. Septembris oli tõesti vaikne, kuid nüüd on küsimisi igasugu- seid."

(12b) ... Septembris oli tõesti vaikne. Nüüd aga on küsimisi igasuguseid.

(12c) Anne Marjapuu, kes on Eesti majandusesindaja Venemaal, kinnitas...

Üheksas tekst (näitetekst 12 ja tulemustekst 12a) demonstreerib otsekõne sisaldava lause lihtsustamist. Vastajad arvasid, et tulemus on tõesti lihtsam, sisu on algsega sama ja lausete järjekord loogiline, samas aga kaotatakse alguses lauses oleva rõhu asukoht. Lisaks pakuti välja lahendusi erinevate kohtade lihtsustamiseks: ühes (12b) tükeldatakse viimane lause otsekõne sees ning teises (12c) soovitati kasutada saatelauses eestäiendi asemel kõrvallauset.

3.4.2. Hinnang muutmatuks jäänud lausetele

Küsitluse teises osas oli välja toodud kolm lauset, mis korpusuuringul jäid silma, sest neid ei lihtsustatud. Samas oli aimdus, et neid oleks võimalik kuidagi lihtsamaks teha. Vastajatel paluti pakkuda välja viis selle teostamiseks.

(13) Samuti toob ülikooli haiglate kokkukolimine vältimatult kaasa Tartu medikute ja abipersonali koondamise, *sest* praegu 15 majas laiali tegutsevad spetsialistid teevad paljusid asju topelt või tegutsevad alakoormusega.

(13a) *Praegu 15 majas laiali tegutsevad spetsialistid teevad paljusid asju topelt või tegutsevad alakoormusega. Seetõttu* toob ülikooli haiglate kokkukolimine vältimatult kaasa ka Tartu medikute ja abipersonali koondamise, *sest*.

(13b) ...*sest* praegu *on* spetsialistid 15 majas laiali, *tehes* paljusid asju topelt või *tegutsedes* alakoormusega.

(13c) *Ülikooli haiglate kokkukolimine toob samuti* vältimatult kaasa...

Esimese lause (13) puhul umbes pooled vastanutest ei lihtsustanud lauset, sest ei osanud või mainimata põhjusel leidsid, et seda ei peaks tegema. Mitmel tekkis probleem algse lause sisust aru saamisega ja välja toodi hoopis küsimust tekitav koht. Kõik lahenduse väljapakunud soovisid teha sellest kaks eraldi lauset, kuid tükelduskoht vastuseti erines. Mitmed leidsid, et sõna *sest* oleks võimalik kaotada ning sealt alustada uut lauset, teised jätsid aga sõna *sest* alustama uut lauset. Arutluse all olevat sõna ka asendati, lause alguses olevana näiteks *Seda sellepärast, et praegu...* või *Seda seetõttu, et praegu...* Keset lauset sooviti seda sõna asendada mõttekriipsuga või sõnaga *kuna*. Kolmel juhul (13a–13c) tõsteti lauses ümber osalused ja muudeti sõnajärge.

(14) Ainult kui Adsonite abielupaar sõitis Pariisi, teatas Pärnu arst, et nüüd neil enam pulbreid ega tilku vaja ei lähe, *vaid* nad võivad kohvikutes ja baarides sisse võtta hoopis parema maitsega mikstuure.

(14a) *Pärnu arst teatas ainult kui* Adsonite abielupaar sõitis Pariisi, et nüüd neil enam pulbreid ega tilku vaja ei lähe. Nad võivad kohvikutes ja baarides sisse võtta hoopis parema maitsega mikstuure.

(14b) Pärnu arst teatas Adsonite abielupaarile, et Pariisis..

(14c) Adsonite abielupaar sõitis Pariisi. *Ainult siis teatas Pärnu arst: "Nüüd teil enam pulbreid ega tilku vaja ei lähe. Selle asemel võite kohvikutes ja baarides sisse võtta hoopis parema maitsega mikstuure."*

Teise lause (14) juures loobus alternatiivse sõnastuse pakkumisest 8 inimest, kellest osad arvasid, et seda ei peaks lihtsustama ning teised soovisid saada laiemat konteksti ja täpsemat infot sisu kohta. Lihtsustamisviisi pakkujatest enamus leidis, et lause oleks sobilik pooleks teha sõna *vaid* juurest. Kohati jäeti see sõna lause algusesse alles, kuid seda asendati seal ka fraasiga *Selle asemel*. Olukorras, kus lauset sealt poolitada ei tahetud, asendati sõna *vaid* näiteks sõnaga *sest* või *ning*. Lahenduste seas oli variante, mida automaatselt on raskem teostada, näiteks fraasi *Ainult kui* asemel alustati tekstilõiku fraasiga *Pärast seda, kui*. Veel oli tõstetud sõna *ainult* lause algusest sõna *arst* järele. Leidus ka lahendusi, kus oli tõstetud fraas *Pärnu arst* lause algusesse (14a, 14b). Sõnavaralihtsustamise kõrval esines ka üks vastus (14c), milles oldi väga julged tegema uusi fraasi- ja lausekonstruktsioone, näiteks tõsteti osa infot otsekõne sisse.

(15) Nagu võis järeltada kaevukoogu ette laotud rohketest pakkidest, kavatses telenaine Põrgupadrikul pikemalt peatuda.

(15a) Kaevukoogu ette laotud rohketest pakkidest *võis järeltada*, et telenaine kavatses Põrgupadrikul pikemalt peatuda.

(15b) *Telenaine kavatses* Põrgupadrikul pikemalt peatuda. *Seda* võis järeltada kaevukoogu ette laotud rohketest pakkidest.

(15c) Telenaise *kavatsusest* Põrgupadrikul pikemalt peatuda *andsid aimu* kaevukoogu ette laotatud *rohked pakid*.

Kolmas (15) muutmatuks jäänud lause näide oli küll lühike, kuid töö autori meelest keeruline kiirelt haarata. Autoriga ei nõustunud 7 inimest, kes vastasid, et nemad seda ei muudaks. Ülejäänud vastused olid vaatamata avatud küsimusele väga üksmeelsed, sest mõlemat tulemust, nii näites 15a kui ka 15b, pakuti välja neli korda. 15a puhul on muudetud mõlemas osalauses sõnajärge ning 15b korral on vahetatud osalausete järjekorda ja tehtud mõlemast eraldi lause. Esines veel variant, kus näites 15b olid laused jäetud kokku üheks tervikuks,

mida ühendas algset lauset alustav sidesõna *nagu*. Automaatseks lahendamiseks raskeks kujuneb aga näites 15c toodud variant, sest seal on muudetud lisaks osalausete järjekorrale ka sõnavorme.

3.4.3. Kasutaja katsetuste tulemused

Küsitluse kolmandas osas sai vastaja jagada katsetatud lauseid ja neid kommenteerida. Lisaks paluti jagada rakenduse kasutuskogemust.

Proovitud tekstid oli väga erisugused, nende hulgas oli nii pikki kui ka ühelauselisi sisendeid, katsetati ka jutumärke, sulge ja erinevaid lausekonstruktsioone sisaldavaid lauseid. Jagatud tekstidest oli näha, et kontrolliti erinevaid teemasid ja päritolusid, sest lauseid mõeldi ise välja, kopeeriti artiklitest, õpikutest või SMS-idest.

Erisugused olid ka programmi tulemused ja sisestajate tagasiside nendele. Kommentaari dena toodi välja mitmeid juhtumeid: programmi väljund oli väga palju segasem kui sisend, tekstis ei muudetud midagi ning oli ka tulemusi, millega jäädi rahule. Sagedasemad kohad, mille puhul vastusega rahul ei oldud, olid tulemusel kolmesõnalised laused, valitud ebasobilikud asesõnad ning kadumaläinud või segasemaks muutunud mõtteedastused. Osades vastustes, millega üldiselt rahule jäeti, esines paaril korral siiski komaviga.

Tagasisides veebirakendusele domineeris mõte, et seda on lihtne ja mugav kasutada, mõned lisasid, et veebileht näeb ilus välja. Neli vastajat mainis, et rakendus töötab kiiresti ning üks tõi välja, et kannatamatu inimese jaoks võib töötlemine küll kaua aega võtta, kuid selliste andmemahtude juures on see normaalne. Küsitluse vormis polnud öeldud midagi sihtgrupi kohta, kuid mitmed pidasid vajalikuks märkida, et sellest võib olla kasu keeleõppijatele, liiga pikkade lausete koostajatele või neile, kes töötavad keerulisemate tekstidega. Avaldati ka imestust, et sellist lihtsustamist saab teha automaatselt.

3.4.4. Kokkuvõtte küsitluse tulemustest

Üldjoontes oli tagasiside rakenduse muudetud lausetele positiivne. Küsitlusest tuli välja, et vastajate hinnangul edastab tulemustekst algsega sama sisu ning asesõnu sisaldavates lausetes oli aru saada, millele pronoomenid viitavad. Vähem üksmeelsed oldi, et muudetud lauses rõhutataks samu kohti. Samas märgiti, et mõnel juhul toob see kogu teksti mõtte paremini esile. Varianti, et tekst muutus lihtsamaks, valiti seda enam, mida pikem oli lause, lühemate lausete puhul ei peetud lihtsustamist vajalikuks. Pikemate lausete lihtsamaks tegemiseks pakuti välja sõnade asendamist ning sõnade või osalausete järjekorra vahetamist.

Kasutajate enda katsetused olid väga mitmekesised, nii sisu kui ka keelekasutuse poolest. Tulemustekstide seas leidis segasemaks muutunud või samasuguseks jäänud tekste ning selliseid, millega jäädi väga rahule. Kasutuskogemuse kohta anti positiivset tagasisidet, muuhulgas toodi välja, et veebileht on kujunduselt ilus ja kasutaja jaoks mõistetav.

3.5. Probleemid

Koostatud programm on reeglipõhine ja seega üritati määratleda võimalikult paljusid juhtumeid. Tekstide loomisel on aga lõputult võtteid ning sarnastes olukordades ei pruugi lahendus olla alati samasugune. Bakalaureusetöö mahu tõttu on selles töös jäänud käsitlemata paljud juhud, millest mõningad on järgnevalt kirjeldatud. Nendest saab lähtuda tehtud programmi edasiarendamisel või uue loomisel.

Praegu esineb mitmeid probleeme ebaselgete pronoomenite asendamisel. Üks variant on see, et valitud arvukategooria on vale. See juhtub näiteks siis, kui leitud viitealune sõna ei viita siiski otsitavale. Töö autorile kui eesti keele emakeelena kõnelejale tundub imelik lause 16 tulemuse (16a) teine lause. Kuigi sõnadega *politsei*, *kiirabi* ja *tuletõrje* kokku kuuluv tegusõna on ainsuses, siis nende ametite esindajad ei käi kunagi üksinda ning uus lause oleks loomulikum algusega *Nad nägid...*

(16) Kohale kihutanud *politsei leidis* mahajäetud mopeedid ja *nägi* minema jooksmas poisse.

(16a) Kohale kihutanud *politsei leidis* mahajäetud mopeedid. *Tema nägi* minema jooksmas poisse.

Eesti keeles on mõningatel asesõnadel lühemad ja pikemad vormid, näiteks *me* ja *meie*. Kuigi neid on võimalik kasutada paralleelselt, siis mõningatel juhtudel võib pikema variandi kasutamine tuleneda soovist tegijat rõhutada. See aspekt tuli välja mitmete näitelause katsetamisel ja ka küsitluse tulemustes, kus oli tunda, et eraldi väljatoodud lihtlause ja seda alustama valitud pronoomen hakkab tekstis tekitama rõhke ja viima seda kontekstis ebavalikku kohta. Näiteks lause 17 tulemuslauses 17a langeb rõhk sõnadele *Mina*.

(17) Olin seisnud hooletult ning kui Piia mind müksas, kaotasin tasakaalu ning langesin kaldalt alla.

(17a) Olin seisnud hooletult, kui Piia mind müksas. *Mina* kaotasin tasakaalu. *Mina* langesin kaldalt alla.

Töö koostamisel tekitasid probleeme osalaised, mis algavad sõnaga, mille lemma on *kes* ja *mis*, sest pealtnäha sarnastes olukordades oli oodatav lahendus täiesti erisugune. Üldiselt oli plaan teha neist eraldiseisvad laused, kuid alati ei olnud see teksti mõistetavuse seisukohast kõige õigem. Näiteks lauses 18 ei saa sõnaga *mis* algavaid osalauseid eraldada ja need tuleks tagasi kokku liita eelmisega. Praegu annab programm tulemuseks näites 18a toodud laused, sest viitealuseks sõnaks loetakse *inimestele* ja seepärast jäetakse järgnevad osalaised eraldiseisvateks. Samas lause 19 puhul tuleks sõnaga *mida* algav osalause kindlasti kokku panna järgmisega.

(18) Olen juba mõnda aega seletanud mujalt maailmast pärit inimestele, *mis* on eesti kunst ja *mis* on temas omapärast ja hinnalist.

(18a) Olen juba mõnda aega seletanud mujalt maailmast pärit inimestele. *Need* on eesti kunst. *Need* on temas omapärast ja hinnalist.

(19) *Mida* loogiliselt ei saa otsustada, tuleb jätta vaistu hooleks.

Lause jälgitavuse seisukohalt on oluline, et oleks aru saada, millele üks või teine pronoomen viitab. Asesõnade automaatsel lisamisel jälgitakse hetkel küll käänat ja võimalusel arvu, kuid siiski tuleb ette olukordi, kus tegija jääb ebaselgeks. Eeltoodud näitelause 2 puhul on praegu programmi tulemuseks tekstilõik *Kass nägi hiirt. Tema kõndis mööda teed. Tema hüppas tema poole*. Võime küll oletada, kellele üks või teine pronoomen *tema* viitab, kuid mõne teise lause puhul ei pruugi see nii läbinähtav olla. Seega võiks tulevikus asendada sõnad *kes* ja *mis* nende viitealustega, mitte pronoomenitega. See tooks lahenduse ka mõningatele eelkirjeldatud probleemidele. Tasub ka meenutada, et peatükis 1.2.2 tutvustatud YATS demo lisas juurde viitava asesõna, mis eesti keeles oleks näiteks *See kass...*

Töös pandi palju rõhku sellele, et leitaks üles õige tegija: võimalusel lause seest või verbi lõpu järgi. Leidub aga juhte, kus tegijat otsima ei peaks või lausa ei tohiks otsida. Näiteks fraasi *tuleb teha* on võimalik tõlgendada nii, et käsk edastatakse kõigile. Kui sinna ette leitakse automaatselt asesõna *Ta*, siis on see ebasobilik. Samuti on olukordi, kus tegija otsimisel leiti õigesse kohta ebasobiv variant. Näiteks lauses 20 ei sobi sõna *leek* asemel öeldise *ronis* ette aluseks *Tema*.

(20) Kiiresti hakkas *leek* limpsima seinapalkide vahelist sammalt ning *ronis* järjest üles poole.

Osalause eraldamisel ja ka tagasi kokku panemisel võib uute sõnade järjekorra tõttu esineda tähendusnihkeid või ebamugavusi loetavuses. Peamiselt tekib olukord siis, kui eraldatud on kiil, mis asus tekstis keset teist süntaktilist üksust. Eraldatud osalause liitmisel eelmisega ei ole aga enam infot, kus see täpselt asus ning tekstilõik lisatakse eelmise osalause lõppu. Lihtlauseks tegemisel tekkiv probleemne koht on näha lauses 21, mille puhul tulemusteksti 21a teine lause võib sellisel kujul niisuguses kohas tekitada suurt segadust. Tagasi kokkupaneku korral asendatakse praeguses tulemusel *Nemad* tekstiga *kes*.

(21) *Need, kes kuidagiviisi müügipunktidesse ei jõua, saavad* pääsme hea õnne korral osta ka mäe pealt enne etenduse algust.

(21a) *Need saavad* pääsme hea õnne korral osta ka mäe pealt enne etenduse algust.
Nemad kuidagiviisi müügipunktidesse ei jõua.

Programmi tulemust mõjutavad suuresti erinevate automaatanalüüside tulemused. Näiteks lauses 22 määratakse sõna *alates* sõnaliigiks verb. Seega hakatakse sellele otsima tegijat ning sõnasid tõstetakse ümber, kuigi ei tohiks (vt näitelause 22a). Analüüsi tulemus omakorda sõltub sisendteksti vormistusest. Kuna osalause tuvastamisel kasutatakse EstNLTK teegis olevat MaltParserit, siis on tekst tundlik õigekirja suhtes, sh komade ja tühikute suhtes. Näitelause 23 juures tuleb tähele panna tühikute olemasolu kriipsu ümber. Kui kasutatud on sidekriipsu, ehk kriips on lühike ja ümbritsetud tühikutega, siis jäetakse kogu lause ühtse tervikuna samasuguseks. Kuid mõttekriipsu esinemisel, ehk kriips on pikk ja ei ole ümbritsetud tühikutega, on tulemuslauseks 23a.

(22) Komisjonil on voli kasutada sõnakuulmatute alaealiste vastu üheksat mõjutusvahendit, *alates* hoiatusest ja lõpetades erikooli suunamisega.

(22a) Komisjonil on voli kasutada sõnakuulmatute alaealiste vastu üheksat mõjutusvahendit. *Suunamisega alates* hoiatusest ja lõpetades erikooli.

(23) Olingi võsast haavametsa piirile välja jõudnud, kui nägin rajaäärsel kivil istuvat naisterahvast - suur õlakott kõrval, tume vihmamantel üle selle heidetud.

(23a) Olingi võsast haavametsa piirile välja jõudnud, kui nägin rajaäärsel kivil istuvat naisterahvast—suur õlakott kõrval. Tume vihmamantel üle selle heidetud.

Veel tekitavad segadust tekstides esinevad arvud ja ka nende kirjutusviis. Lauses 24 on probleemne fraas *12 000-kroonine*. Viimaseks eraldi osalauseks analüüsitakse tekst *alates* sõnast *ning* ja sealt otsitakse verbile *ähvardama* tegijat. Analüüside tulemusel leitakse tegijaks tühiku korral arv *12* (vt lause 24a) või tühikuta kirjutamisel *12000-kroonine* (vt lause

24b) ja tuuakse see lause ette. Tulemust ei saa kummalgi juhul lugeda korrektseks. Korpus-tekstis oligi see kirjutatud näidatud viisil tühikuga, kuid eestikeelsetes tekstides esineb ka võõrapärasusi ja tuhandikke eraldatakse punkti või komaga.

(24) Tallinnas elavad sõidukiomanikud peavad tasuma selle aasta mootorsõidukimaksu laupäevaks, maksmata jätmisel lisandub iga päev viivis ning ähvardab üle 12 000-kroonine trahv.

(24a) [---] 12 ähvardab üle 000-kroonine trahv.

(24b) [---] 12000-kroonine trahv ähvardab üle.

Mitmeid lahendusi kaaluti sulge sisaldava lause jaoks. Üldiselt ei märgita sulgudesse terveid lausekatkeid, vaid täpsustavaid infokilde. Seetõttu oli mõte kogu sulgudes olev tekst tulemusest kaotada. Teiseks võimaluseks peeti seda, et kui sulgudes olevate sõnade hulgas esineb verb, siis teha sellest omaette lause. Kuna valitud korpus oli sulge sisaldavaid näiteid küllaltki vähe, siis ei osanud töö autor hinnata ühe või teise tegevusstenaariumi mõju ning lõpptulemusena säilitati sulgudes olev info ja selle asukoht võimalikult algse lähedaselt.

3.6. Edasiarendusvõimalused

Selles töös keskenduti lausestruktuuri lihtsustamisele, et liitlausetest saaksid lihtlauseid. Nagu Shardlow (2014: 62) oma artiklis välja tõi, saab pöörata tähelepanu ka teistele süntaksi alla kuuluvatele aspektidele. Näiteks saaks lihtsustada ajavormide kasutust ning eraldi lauseks moodustada sekundaartarindid. Viimased on sündmust väljendavad moodustajad, aga pole alati omaette osalused, näiteks *Lõputööd kirjutades kontakteerus ta juhendajatega*. See eeldaks aga suuremat muutmist kogu algse lause ulatuses. Veel saaks muuta *poolt*-tarindeid. Sellest johtuvalt võiks kaaluda Kaarel Sõrmuse (2017) loodud kantseliidi- ja paronüümituvastaja kasutamist lihtsustamise programmi tulemuses, sest seal on käsitletud ka teisi teksti selguse parandamiseks antud soovitusi.

Kuigi leksikaalne ja süntakiline lihtsustamine on kasulikud veidi erinevatele sihtrühmadele, siis leidub ka neid, kes vajaksid tekstide puhul mõlemat lihtsustamist korraga. Seega võiks ühendada Martin Peedoski loodud sõnavaralihtsustamise programmi (2017a) lauseehituse lihtsustajaga ühtsesse rakendusse.

Siinses töös olev lähenemine lihtsustamisele on reeglipõhine ehk on etteantud juhud ja nende lahendused. Inglise keelele on aga hakatud arendama masinõppel põhinevat lihtsustamisprogrammi, mis kohandab nii sõnavara kui ka lausestust (Coster ja Kauchak, 2011). Selle projekti tegemist lihtsustab muidugi asjaolu, et inglise keeles on olemas kaks paralleelset Vikipeediat, millest ühes on kasutatud nii-öelda lihtsamat keelt. Siiski võib loota, et kunagi on sarnased paralleelkorpustena kasutatavad tekstid olemas ka eesti keeles ning teksti lihtsustamist saab teha masinõppe abil.

Kokkuvõte

Selle bakalaureusetöö eesmärk oli luua reeglipõhine eestikeelse teksti lauseehitust lihtsustav programm. See võiks aidata näiteks keeleõppijaid ja keskendumishäiretega inimesi, sest lihtlause puhul saab reeglina korruga keskenduda vaid ühele mõttele. Leidmaks parimaid võimalusi lihtsustaja loomiseks, tutvuti sarnaste töödega, mis on tehtud teistele keeltele.

Töös tutvustati erinevaid teksti lihtsustamise viise, mida võib ka omavahel kombineerida. Leksikaalse ehk sõnavara lihtsustamise korral asendatakse tekstis vähemlevinud ja pikemad sõnad lühemate ja sagedasemate sõnadega. Süntaktilise ehk lausestuse lihtsustamise puhul saab keskenduda erinevatele lausetüüpidele, mõnele grammatilisele kategooriale või teatud tüüpi tarinditele. Samuti saab teksti lihtsustada seda kokku võttes, näiteks tuues välja olulisemad laused või tähtsamad märksõnad.

Teksti lihtsustamise vastu on huvi tuntud juba väga mitme keele puhul. Süntaktilisest lihtsustamisest tehtud uurimusi tutvustati ja võrreldi järgnevate keelte kohta: Brasiilia portugali, itaalia, baski, vietnami ja korea. Need kõik olid reeglipõhised lähenemised, kus võeti aluseks tuvastatud osalaused või kindlat tüüpi fraasid ja täiendid. Lisaks eraldi lausete tegemisele pöörati näiteks portugali keele puhul tähelepanu ka *poolt*-tarinditele ning itaalia keele korral jäeti alles vaid faktilist infot edastavad laused, mis muudeti olevikku. Korea keelele mõeldud rakendus aga visualiseerib tekstis osalaused ja nende omavahelised suhted.

Täpsemalt kirjeldati inglise keelega katsetatavat veebirakendust YATS demo, mis on algselt loodud hispaania keelele. Selle puhul analüüsiti komplekslause ja otsekõnet sisaldava lause lihtsustamist, komavigade tundlikkust ning muuhulgas tuli välja ka sõnavara lihtsustamise funktsioon. Kasutamiskogemusest võeti tehtava rakenduse veebikujunduse jaoks eeskujuks lehekülje lihtne ülesehitus.

Töö praktilise osana valminud eestikeelse teksti lausestruktuuri lihtsustav programm on kasutatav veebirakendusena Eesti Keeleressursside Keskuse serveris. Selle loomisel toetuti eesti keele analüüsimiseks mõeldud tehnilistele vahenditele, eelkõige osalausestajale, kuid selle kõrval ka süntaksianalüsaatorile, morfoloogilisele analüsaatorile ja morfoloogilisele süntesaatorile. Kõiki neid ka tutvustati põhjalikumalt töö kirjalikus osas.

Lihtsustamisprogrammi tulemustekstis üritatakse säilitada teksti algne struktuur lõikude kaupa ning sulge, tsitaati või pealkirja sisaldava lause puhul ka selle asukohta osalauses.

Seega analüüsitakse kõigepealt jutumärkide olemasolu, seejärel kooloni ning viimasena sulgude olemasolu lauses. Otsekõnet sisaldava lause puhul viiakse võimalusel saatelause otsekõne ette, kujule *X ütles: „...“*. Üldiselt üritatakse igast osalausest teha eraldi lihtlause, kuid teatud olukordades ei ole see teksti selguse huvides võimalik.

Loodud programmile hinnangu andmiseks tehti küsitlus, sest keelekasutuse lihtsust ei saa hinnata valemi või reeglistiku abil, vaid subjektiivselt. Igapäevase eesti suhtlemis- ja lugemiskeelega 20 inimest tagasisidestasi üheksat muudetud lauset, kolme muutmatuks jäänud lauset ning seejärel said nad ise veebirakendust ka katsetada. Kuigi kõiki tulemustekste ei hinnatud algsest lihtsamaks, pakuti selle kõrvale välja mitmeid lahendusi. Mitmete näidete puhul toodi välja, et rõhuasetuse muutus või tulemuse vastuvõetamatus on tingitud asesõna pika vormi (*Tema*) kasutamisest lühikese (*Ta*) asemel. Samuti toodi välja, et väga lühikesi lauseid pole mõtet tükeldada. Üldiselt leiti aga, et tulemustekst edastab sama sisu. Veebirakendusele endale anti positiivset tagasisidet, sest programmi on lihtne ja mugav kasutada.

Kuna tekstide loomisel on väga palju erinevaid võimalusi, siis neid kõiki ei saa bakalaureusetöö mahu tõttu käsitleda ega rakenduse loomisel teostada. Kirjalikus osas analüüsiti mõningaid silmatorkavamaid probleeme tulemustekstides. Suuremad murekohad olid seoses pronoomenite asendamisega ja õigete tegijate leidmisega, samuti juhtudega, kus pronoomeniga algav osalause tuleks liita eelmise või järgmisega. Mõningatel juhtudel segasid õige tulemuse saavutamist teiste tehniliste vahendite antud analüüsid.

Lisaks praeguse rakenduse tulemuses levivate vigade parandamisele võiks edaspidi pöörata tähelepanu ka muudele lihtsustamisaspektidele. Teiste keelte eeskujul võiks visualiseerida osalauseid ja lihtsustada teisi süntaksi alla kuuluvaid tegureid. Samuti võiks parandada paronüümset ja kantseliitlikku keelekasutust ning ühendada süntaktiline ja leksikaalne lihtsustaja ühte rakendusse. Vastupidiselt siin väljatoodud reeglipõhisele lähenemisele võiks aga ka teha hoopis masinõppel põhineva teksti lihtsustaja.

Viidatud kirjandus

- Aluísio, S. M. & Gasperin, C. (2010). Fostering Digital Inclusion and Accessibility: The PorSimples Project for Simplification of Portuguese Texts. *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas* (p. 46–53). Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=1868701.1868708> (vaadatud 13.03.2019).
- Aranzabe, M. J., Diaz de Ilarraza, A. & Gonzalez-Dios, I. (2012). First Approach to Automatic Text Simplification in Basque. *Natural Language Processing for Improving Textual Accessibility (NLP4ITA)*, 1–8. <http://www.lrec-conf.org/proceedings/lrec2012/workshops/25.NLP4ITA-Proceedings.pdf#page=6> (vaadatud 13.03.2019).
- Aranzabe, M. J., Ilarraza, A. D. de & Gonzalez-Dios, I. (2013). Transforming Complex Sentences using Dependency Trees for Automatic Text Simplification in Basque. *Procesamiento del Lenguaje Natural*, 50, 61–68. <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/4660/2762> (vaadatud 13.03.2019).
- Barlacchi, G. & Tonelli, S. (2013). ERNESTA: A Sentence Simplification Tool for Children's Stories in Italian. A. Gelbukh (Eds.), *Computational Linguistics and Intelligent Text Processing. 14th International Conference, CICLing 2013, Samos, Greece, March 24–30, 2013, Proceedings, Part II*. (p. 476–487). Samos, Greece: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37256-8_39
- Chandrasekar, R., Doran, C. & Srinivas, B. (1996). Motivations and Methods for Text Simplification. *Proceedings of the 16th Conference on Computational Linguistics - Volume 2* (p. 1041–1044). Stroudsburg, PA, USA: Association for Computational Linguistics. <https://doi.org/10.3115/993268.993361>
- Chung, J.-W., Min, H.-J., Kim, J. & Park, J. C. (2013). Enhancing readability of web documents by text augmentation for deaf people. *Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics - WIMS '13*, 1. New York, USA: ACM Press. <https://doi.org/10.1145/2479787.2479808>

- Coster, W. & Kauchak, D. (2011). Learning to Simplify Sentences Using Wikipedia. *Proceedings of the Workshop on Monolingual Text-To-Text Generation (MTTG '11)* (p. 1–9). Stroudsburg, PA, USA: Association for Computational Linguistics. <http://dl.acm.org/citation.cfm?id=2107679.2107680> (vaadatud 11.04.2019).
- Daelemans, W., Höthker, A. & Sang, E. T. K. (2004). Automatic Sentence Simplification for Subtitling in Dutch and English. *Proceedings of the 4th International Conference on Language Resources and Evaluation*, 1045–1048. <http://lrec-conf.org/proceedings/lrec2004/pdf/697.pdf> (vaadatud 13.03.2019).
- Eesti keele süntaksianalüsaator. (i.a.). <http://kodu.ut.ee/~kaili/parser/> (vaadatud 29.12.2018).
- Eesti keele välisõpe. (i.a.). *Haridus- ja Teadusministeerium*. <https://www.hm.ee/et/tegevused/keeled/eesti-keele-valisope> (vaadatud 28.12.2018).
- Eesti keele õpe välismaal. (i.a.). *Eesti Keel ja Kultuur Maailmas*. <https://ekkm.estinst.ee/eesti-keele-ope-valismaal/> (vaadatud 28.12.2018).
- Erelt, M., Erelt, T. ja Ross, K. (2007). *Eesti keele käsiraamat* (Kolmas, täiendatud trükk). Tallinn: Eesti Keele Sihtasutus.
- EstNLTK dokumentatsioon*. (i.a.). <https://estnlk.github.io/> (vaadatud 29.12.2018).
- Ferrés, D., Marimon, M., Saggion, H. & AbuRa'ed, A. (2016). YATS: Yet Another Text Simplifier. E. Métais, F. Meziane, M. Saraee, V. Sugumaran ja S. Vadera (Eds.), *Natural Language Processing and Information Systems. 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings*. (p. 335–342). Salford, UK: Springer International Publishing. http://doi.org/10.1007/978-3-319-41754-7_32
- Filosoft*. (i.a.). <http://www.filosoft.ee/> (vaadatud 29.12.2018).
- Freienthal, L. (2018). *Pronominaalsete viitesuhete analüüs asendussõnade suhtes käsitsi märgendatud korpuses*. Bakalaureusetöö. Tartu Ülikool, eesti ja üldkeeleteaduse osakond. <https://dspace.ut.ee/handle/10062/60373> (vaadatud 23.03.2019).
- Gao, J. ja Shen, D. (2018). Eesti keele põhierialast Hiinas. *Oma Keel*, 2, 64–69.

- Hallik, K. (2011). Selge keele liikumisest maailmas. *Õiguskeel*, 4.
https://www.just.ee/sites/www.just.ee/files/katrin_hallik._selge_keele_liikumisest_maailmas.pdf (vaadatud 27.12.2018).
- Hung, B. T., Minh, N. L. & Shimazu, A. (2012). Sentence Splitting for Vietnamese-English Machine Translation. *2012 Fourth International Conference on Knowledge and Systems Engineering (KSE)*, 156–160. <https://doi.org/10.1109/KSE.2012.28>
- Kaalep, H.-J. (1998). Tekstikorpuse abil loodud eesti keele morfoloogiaanalüsaator. *Keel ja Kirjandus*, 1, 22–29.
- Kaalep, H.-J. ja Muischnek, K. (2012). Osalausete tuvastamine eestikeelses tekstis kui iseseisev ülesanne. *Eesti Rakenduslingvistika Ühingu Aastaraamat*, 8, 55–68.
<https://doi.org/10.5128/ERYa8.04>
- Kaalep, H.-J. ja Vaino, T. (1998). Kas vale meetodiga õiged tulemused? Statistikaline tugi- nev eesti keele morfoloogiline ühestamine. *Keel Ja Kirjandus*, 1, 30–38.
- Laur, S. (i.a.). EKT57 „EstNLTK: Pythoni teegid eestikeelsete vabatekstide lihtsamaks töötlemiseks (1.01.2014–31.12.2016)“. *ETIS*. <https://www.etis.ee/Portal/Projects/Display/0b4c5efa-6b2c-40cd-aff6-4a6b6668560a> (vaadatud 29.12.2018).
- Mikk, J. (2015, november 20). Õppematerjalide jõukohasuse elektrooniline hindamine. *Õpetajate Leht*. <http://opleht.ee/2015/11/oppematerjalide-joukohasuse-elektrooniline-hindamine/> (vaadatud 28.12.2018).
- Milne, A. A. (1981). *Winnie-the-Pooh*. Great Britain: Richard Clay Ltd.
- Müürisep, K. (1996). *Eesti keele kitsenduste grammatika süntaksianalüsaator*. Magistri- töö. Tartu Ülikool, arvutiteaduse instituut.
- Müürisep, K. (2000). *Eesti keele arvutigrammatika: süntaks*. *Dissertationes Mathematicae Universitatis Tartuensis* 22. Tartu: Tartu Ülikooli kirjastus.

- Müürisep, K. (2004). Eesti keele süntaksianalüsaatori märgenditest. H. Metslang (toim); M. Sepper, J. Lepasaar (koost). *Toimiv keel. II: töid rakenduslingvistika alalt: Tallinna Pedagoogikaülikooli eesti filoloogia osakonna toimetised 3. Tallinnas Pedagoogikaülikool, Eesti Keele Instituut* (lk 207–217). Tallinn: Tallinna Pedagoogikaülikooli Kirjastus. <http://kodu.ut.ee/~kaili/papers/myyrisep prakling03final.pdf> (vaadatud 29.12.2018)
- Müürisep, K. (2006). Eestikeelsete tekstide sisukokkuvõtjast EstSum. *Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6*, 115–125.
- Müürisep, K., Nigol, H. ja Uibo, H. (2006). Eesti suulise keele korpuse automaatne pind-süntaktiline analüüs. *Keel ja arvuti. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 6*, 72–84.
- Pajusalu, K. (2018). *Suuline teade loengukursusel „Eesti keele ajalugu“*. Tartu, november.
- Peedosk, M. (2017a). *Eesti keele digitaalsete ressursside ja tehnoloogiate rakendamine teksti lihtsustamise programmis*. Bakalaureusetöö. Tartu Ülikool, arvutiteaduse instituut. https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=58269&jayear=2017 (vaadatud 28.12.2018).
- Peedosk, M. (2017b). *Teksti lihtsustamine*. <http://prog.keeleressursid.ee:4567/> (vaadatud 8.04.2019).
- Raadik, M. (2014). Mis kujuga on jutumärgid? *Oma Keel*, 2, 49–50.
- Rybing, J., Smith, C. & Silvervarg, A. (2010). Towards a Rule Based System for Automatic Simplification of Texts. *The Third Swedish Language Technology Conference (SLTC 2010)*, 17–18. <https://www.ida.liu.se/conferences/sltc2010/sltc2010-proceedings.pdf#page=25> (vaadatud 28.12.2018).
- Saggion, H., Štajner, S., Bott, S., Mille, S., Rello, L. & Drndarevic, B. (2015). Making It Simplext: Implementation and Evaluation of a Text Simplification System for Spanish. *ACM Transactions on Accessible Computing (TACCESS) - Special Issue on Speech and Language Processing for AT (Part 2)*, 6(4), 1–36. <https://doi.org/10.1145/2738046>

- Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Natural Language Processing 2014*, 58–70. <https://doi.org/10.14569/SpecialIssue.2014.040109>
- Siddharthan, A. (2006). Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4(1), 77–109. <https://doi.org/10.1007/s11168-006-9011-1>
- Simplext – Text simplifier. (2016). *Able to Include*. <http://able-to-include.com/accessible-services/services/text-to-easy-to-read/> (vaadatud 14.03.2019).
- Sõrmus, K. (2017). *Kantseliidi- ja paronüümituvastaja*. Bakalaureusetöö. Tartu Ülikool, arvutiteaduse instituut. https://comserv.cs.ut.ee/ati_thesis/datasheet.php?id=58321&year=2017 (vaadatud 28.12.2018).
- YATS demo. (2016). <http://able2include.taln.upf.edu/> (vaadatud 14.03.2019).
- Yle Uutiset selkosuomeksi. (s.a.). <https://yle.fi/uutiset/osasto/selkouutiset/> (vaadatud 28.12.2018).

Lisad

I. Küsitlus

Lauseehituse lihtsustamine

Küsitlusega kogutakse tagasisidet Ruth Schihalejevi bakalaureusetööna valmivale programmile, mis lihtsustab teksti lauseehitust (st muudab liitlause lihtlauseteks, jättes mõtte samaks). Küsitlus koosneb neljast osast, esimeses osas küsitakse hinnangut programmi tulemusel muutunud lausetele; teises muutumatuks jäänud lausetele; kolmandas saab osaleja tagasisidestada enda proovitud lauseid ja veebirakenduse kasutust üldiselt ning neljandas kogutakse mõningad osaleja taustaandmed.

Küsitlus on anonüümne ja vabatahtlik, selles osalemise võib igal ajahetkel katkestada. Saates vastused ära, annab osaleja nõusoleku vastuseid eelmainitud töös kasutada.

Tulemustega on võimalik tutvuda 2019. suvel kaitstavas bakalaureusetöös.

Kontaktisik: Ruth Schihalejev

Küsitluses kasutatud näited on pärit bakalaureusetöö jaoks kitsendatud Tasakaalus korpusest, kasutatud on nii ilukirjandus- kui ka ajakirjandustekste. Tekstid on leitavad aadressilt

<https://github.com/ruthsch/lauseehituse-lihtsustamine/tree/master/korpusfailid>.

JÄRGMINE

Ärge saatke paroole kunagi Google'i vormide kaudu.

Lauseehituse lihtsustamine

* Kohustuslik

Etteantud muutunud tekstide hindamine

Järgnevalt on esitatud lausete paare, millest esimene on algne tekstilõik (või lause) ning teine programmi koostatud tekst. Palun hinnata, kas muutunud teksti on kergem lugeda? Kas algset teksti võiks kuidagi teisiti lihtsustada, kasutades seal olevaid elemente, sõnu ja infot?

Tekst 1

ALGNE:

Naisel purunes kulm ja sai viga nina, kaasa sõitnud autoomanikust noor mees jäi terveks.

MUUDETUD:

Naisel purunes kulm. Viga sai nina. Kaasa sõitnud autoomanikust noor mees jäi terveks.

Vali Sinu hinnangul kõik tulemusteksti nr 1 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 2

ALGNE:

Aga härra Kallion oli alles noor jurist ning ihaldas kiiresti saavutada pealinna üleviimist.

MUUDETUD:

Aga härra Kallion oli alles noor jurist. Tema ihaldas kiiresti saavutada pealinna üleviimist.

Vali Sinu hinnangul kõik tulemusteksti nr 2 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- On aru saada, kellele/millele asesõna(d) viitab(/-vad)
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 3

ALGNE:

Kõiki müügikohti haldab ainult üks omanik, mis tähendab, et praeguses ühiskonnas enesestmõistetav vaba konkurents Elva linna kõige lootustandvamal äripinnal ei toimi.

MUUDETUD:

Kõiki müügikohti haldab ainult üks omanik. See tähendab, et praeguses ühiskonnas enesestmõistetav vaba konkurents Elva linna kõige lootustandvamal äripinnal ei toimi.

Vali Sinu hinnangul kõik tulemusteksti nr 3 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- On aru saada, kellele/millele asesõna(d) viitab(/-vad)
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

[Teie vastus](#)

Tekst 4

ALGNE:

Kuna need, kes pomme panevad, ei ähvarda, ja need, kes ähvardavad, ei pommita, siis näib pommikoerte jooksutamine iga idioodi telefonikõne peale mõttetu tegevusena.

MUUDETUD:

Kuna need ei ähvarda, kes pomme panevad, ja need ei pommita, kes ähvardavad, siis näib pommikoerte jooksutamine iga idioodi telefonikõne peale mõttetu tegevusena.

Vali Sinu hinnangul kõik tulemusteksti nr 4 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- On aru saada, kellele/millele asesõna(d) viitab(/-vad)
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 5

ALGNE:

Olin juba ligi aasta vabakutseline ning pidin rohkem ridasid paberile nõiduma kui palgalised ajakirjanikud, kes võisid endale vaheldust leida baarides istumise ja komanderingutes käimisega.

MUUDETUD:

Olin juba ligi aasta vabakutseline. Mina pidin rohkem ridasid paberile nõiduma kui palgalised ajakirjanikud. Nemad võisid endale vaheldust leida baarides istumise ja komanderingutes käimisega.

Vali Sinu hinnangul kõik tulemusteksti nr 5 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- On aru saada, kellele/millele asesõna(d) viitab(/-vad)
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Tekst 6

ALGNE:

Tarmo Kahera, kes lesis kollase kambri diivanil televiisori vastas, ootas mult vastust.

MUUDETUD:

Tarmo Kahera ootas mult vastust. Tema lesis kollase kambri diivanil televiisori vastas.

Vali Sinu hinnangul kõik tulemusteksti nr 6 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- On aru saada, kellele/millele asesõna(d) viitab(/-vad)
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 7

ALGNE:

Wallingatan meenutab kangesti Tallinna vanalinna, tundus, nagu asuksin Rataskaevu ja Pika tänava nurgal.

MUUDETUD:

Wallingatan meenutab kangesti Tallinna vanalinna ja tundus, nagu asuksin Rataskaevu ja Pika tänava nurgal.

Vali Sinu hinnangul kõik tulemusteksti nr 7 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 8

ALGNE:

Oli vaja kraavikaldalt metsa keerata - minu arusaamist mööda ei viinud veenire talu, vaid jõe poole.

MUUDETUD:

Oli vaja kraavikaldalt metsa keerata. Minu arusaamist mööda ei viinud veenire talu, vaid jõe poole.

Vali Sinu hinnangul kõik tulemusteksti nr 8 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Tekst 9

ALGNE:

"Huvi meie kauba vastu tõuseb," kinnitas Eesti majandusesindaja Venemaal Anne Marjapuu. "Septembris oli tõesti vaikne, kuid nüüd on küsimisi igasuguseid."

MUUDETUD:

Eesti majandusesindaja Venemaal Anne Marjapuu kinnitas: "Huvi meie kauba vastu tõuseb. Septembris oli tõesti vaikne, kuid nüüd on küsimisi igasuguseid."

Vali Sinu hinnangul kõik tulemusteksti nr 9 kohta käivad tõesed väited: *

- On algsest lihtsam
- Sisu on algsega sama
- Rõhutatakse algse tekstiga sama kohta
- Sõnajärg on loogiline
- Lausete järjekord on loogiline
- Kõik laused on eesti keele grammatikale vastavad
- Ükski väide ei kehti

Vajadusel lisa kommentaare ja/või paku parem viis lauseehituse lihtsustamiseks:

Teie vastus

Etteantud muutumatuks jäänud tekstide hindamine

Järgnevalt on esitatud laused, mida programm ei muutnud. Kas peaks neid siiski kuidagi lihtsustama? Kui jah, siis kuidas võiks teksti lihtsustada, kasutades seal olevaid elemente, sõnu ja infot? Vajadusel võib muuta sidesõnu.

Tekst 1

Samuti toob ülikooli haiglate kokkukolimine vältimatult kaasa Tartu meedikute ja abipersonali koondamise, sest praegu 15 majas laiali tegutsevad spetsialistid teevad paljusid asju topelt või tegutsevad alakoormusega.

Paku, kuidas lihtsustada lauseehitust ja/või lisa kommentaare: *

Teie vastus

Tekst 2

Ainult kui Adsonite abielupaar sõitis Pariisi, teatas Pärnu arst, et nüüd neil enam pulbreid ega tillku vaja ei lähe, vaid nad võivad kohvikutes ja baarides sisse võtta hoopis parema maitsega mikstuure.

Paku, kuidas lihtsustada lauseehitust ja/või lisa kommentaare: *

Teie vastus

Tekst 3

Nagu võis järeldada kaevukoogu ette laotud rohketest pakkidest, kavatses telenaine Põrgupadrikul pikemalt peatuda.

Paku, kuidas lihtsustada lauseehitust ja/või lisa kommentaare: *

Teie vastus

Kasutaja katsetus(t)e tulemused

Palun katseta rakendust aadressil <http://prog.keeleressursid.ee/lihtlauseks/>. Tekst võib pärineda näiteks mõnest artiklist ning sisendkasti on võimalik lisada korraga ka mitu lõiku. Kommenteeri siin oma katselausete tulemusi ja/või üldist rakenduse kasutuskogemust.

Lisa siia katsetatud sisendtekst(id): *

Teie vastus

Kommenteeri sisendteksti(de) tulemust *

Teie vastus

Tagasiside veebirakenduse kasutamisele: *

Teie vastus

Vastaja taustaandmed

Sisesta oma vanus *

Valige ▼

Kas eesti keel on Sinu igapäevane suhtlemis- ja lugemiskeel? *

Jah

Ei

TAGASI

SAADA ÄRA

Ärge saatke paroole kunagi Google'i vormide kaudu.

II. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Ruth Schihalejev**,

(autori nimi)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
„Reeglipõhine lausekonstruktsioonide lihtsustaja eestikeelsetes tekstides“,
(lõputöö pealkiri)

mille juhendajad on Sven Aller ja Heili Orav,

(juhendaja nimi)

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Tartus, **09.05.2019**