TARTU ÜLIKOOL
Loodus- ja täppisteaduste valdkond
Arvutiteaduse instituut
Informaatika õppekava

Anneliis Halling

Eesti keele keeleressursse kasutav õppeprogramm käänete õppimiseks

Bakalaureusetöö (9 EAP)

Juhendaja: Sven Aller

Tartu 2016

Eesti keele keeleressursse kasutav õppeprogramm käänete õppimiseks

Lühikokkuvõte:

Käesoleva bakalaureusetöö eesmärk oli luua veebipõhine õppeprogramm eesti keele käänete õppimiseks. Eesti keele käänete õppimise programm peaks sobima nii põhikoolis kui ka gümnaasiumis õppivatele õpilastele ja ka mitte-eestlastele, kes soovivad õppida eesti keelt süvendatult. Õppeprogrammi loomiseks kasutati koos esmakordselt selliseid keeleressursse nagu morfoloogiline analüsaator ja süntesaator, ilukirjanduskorpus, sagedussõnastik ning ebasobivate sõnade loend. Eesmärgi täitmiseks loodi reeglid, et valida välja õppeprogrammi jaoks sobivad laused. Töös kirjeldatakse nii eeltöötlusprogrammi kui ka õppeprogrammi algoritmi.

Võtmesõnad: õppeprogramm, tekstikorpus, keeleressursid, eesti keele käänded

CERCS: P175

A program for learning cases by using Estonian language resources

Abstract:

The aim of this Bachelor's thesis was to create a web-based program for learning Estonian cases. This learning program should fit from elementary school to high school pupils, and also for foreigners, who want to learn the Estonian language in depth. During the creation of this program for learning cases, some of the Estonian language resources were used together for the first time, such as the morphological analyzer and synthesizer, fiction corpus, frequency dictionary and a list of inappropriate words. In order to achieve this, some rules were made to sort out suitable sentences. The thesis also includes the explanation of the pre-processing algorithm and the learning program algorithm.

Keywords: learning program, corpus-based learning, language resources, Estonian cases

CERCS: P175

Sisukord

Sissejuhatus

Keel on üks tähtsamaid eesti rahvuse tunnusjooni ja meie kultuuri kandja. Keeleõpe on Eestis väga tähtsal kohal, kuna eesti keelt kõnelevaid inimesi on vaid 1,25 miljonit [1] ja see arv väheneb [2]. Eesti keel kuulub soome-ugri keelte hulka ning neid keeli peetakse keerukateks muuhulgas ka käänete rohkuse tõttu. Võrreldes inglise keelega tuleb eesti keele grammatikat ja sõnavara põhjalikumalt õppida, enne kui saab eesti keeles korrektselt rääkida [3] ning seetõttu tekitab grammatika õppimine nii välismaalaste kui ka eestlaste endi seas raskusi. Eesti keeleõppes on alati teretulnud täiendav õppematerjal, millega saaks põhjalikult ja lihtsalt õppida grammatikat, sealhulgas käändeid.

Käändeid õpitakse nii põhikoolis kui ka gümnaasiumis. Nende õppimiseks ja harjutamiseks on mitmeid võimalusi. Näiteks on võimalik lahendada interaktiivseid töölehti, veebirakendustes teste ja grammatikat õpetavates töövihikutes erinevaid käänete ülesandeid. Sellist eesti keele käänete õppeprogrammi ei ole loodud, millega saaks õppida käändeid põhjalikult ning mis sisaldaks palju mitmesuguseid näitelauseid. Keeleõppes kasutatakse üle kogu maailma aina laialdasemalt tekstikorpuseid, kuna need pakuvad suurel hulgal ehedaid ja mitmekesiseid lauseid [4]. Seega sobivad tekstikorpuse laused hästi käänete õppeprogrammi loomiseks.

Käesolev bakalaureusetöö on rakenduslik uurimustöö ning selle eesmärk on keeleressursse kasutades luua veebipõhine õppeprogramm eesti keele käänete õppimiseks. Antud õppeprogramm peaks sobima nii põhikoolis kui ka gümnaasiumis õppivatele õpilastele ja ka mitte-eestlastele, kes soovivad õppida eesti keelt süvendatult. Õppeprogrammiks tarvilikud laused saadakse eesti keele tekstikorpusest. Eesmärgi saavutamiseks kasutatakse ka morfoloogilist analüsaatorit ja süntesaatorit, sagedussõnastikku ning ebasobivate sõnade loendit, et välja valida keeleõppe jaoks sobivad laused.

Töö on jagatud kolmeks peatükiks. Esimeses peatükis kirjeldatakse ja võrreldakse olemasolevaid õppeprogramme, mis on loodud eesti keele käänete õppimiseks. Teises peatükis uuritakse ning tutvustatakse keeleressursse, mida kasutatakse õppeprogrammi loomisel. Kolmandas peatükis kirjeldatakse valminud õppeprogrammi, selle algoritmi, tekkinud probleeme ning edasiarendamise võimalusi.

1. Eesti keele käänete õppeprogrammid ja interaktiivsed testid

Tehnoloogia areng ja internet võimaldavad luua palju mitmekesisemaid õppematerjale kui varem ning kasutada uusi õppemeetodeid [5]. Üheks võimaluseks on õppeprogrammid, mis pakuvad head vaheldust õpikute ja töövihikute ülesannetele ning sobivad nii noortele kui vanadele. Kui õppeprogrammide materjal on mitmekesine, eluliste näidetega ning olemas on tagasiside saamise võimalus, siis muudab see õppimise huvitavaks ja tõhusaks [5]. Seega saab õppeprogrammide abil väga hästi õpetada grammatikat. Grammatika õpetamiseks on mitmeid meetodeid. Ingrid Krall ja Elle Sõrmus [3] on kirjutanud, et induktiivne õppimine võimaldab enne grammatika reeglite selgitamist tunnetada keelt ning nendega õppimise käigus tutvuda. Nad usuvad, et selline lähenemine lisab motivatsiooni edasi õppimiseks ja avastamiseks. Käänete õppimise kontekstis on induktiivse õppimise heaks näiteks lausest eemaldatud sõna õigesse käändesse panemine. Induktiivse õppimise kõrval saab varem õpitut kontrollida viisil, kus ette on antud sõna ning tuleb määrata kääne. Selliseid ülesandeid saab väga hästi rakendada õppeprogrammides. Lisaks eelöeldule seisneb veebipõhiste õppeprogrammide eelis selles, et nende puhul on võimalus anda kohest tagasisidet. Samuti on võimalik kõiki andmeid salvestada, millest hiljem saab teha statistilisi järeldusi – näiteks, millised käänded on paremini või halvemini omandatud. Kohene tagasiside on kasulik, kuna kasutaja saab oma vigadest õppida pärast igat vastuse esitamist.

Käänete õppimiseks leidub mitmesuguseid teste ja rakendusi, mis on mõeldud eelkõige lahendamiseks käänetega esmakordsel tutvumisel ning seega jäävad gümnasistidele liiga lihtsaks. Valdavalt on rakendustes ülesanded, kus tuleb määrata käände küsimus, kirjutada sõna õigesse käändesse (koos lausega ja ilma) või määrata sõna kääne.

1.1 Valmisvahenditega loodud interaktiivsed testid

Ülesannete loomise tarkvaraga Hot Potatoes [6] on tehtud käänete õppimiseks interaktiivseid teste. Selliseid teste leiab näiteks veebilehtedelt Sahver [7] ja Eesti keel [8].

Lehel Sahver [7] on kuus testi 5. klassile. Testid tuleb alla laadida HTML (*HyperText Markup Language*) failidena ning avada brauseris.

Teste on kahte tüüpi:

1. „Esita sõna kohta küsimus“, kus tuleb sõna järgi kirjutada küsisõna, millele sõna vastab.
2. „Harjuta sõna kohta esitatud küsimuse järgi käände ja arvu määramist“, kus tuleb sõna järgi kirjutada kääne ja arv, milles on sõna.

Kõik testid on esitatud lausete kontekstis ning enamus testide lauseid on võetud ilukirjandusest. Valesti vastatud küsimustele saab uuesti vastata.

Eesti keele veebilehel [8] on seitse interaktiivset käänete testi, kus tuleb:

1. kirjutada lünka antud sõna nõutud arvus ja käändes;
2. leida vasakul olevale sõnale valikuribalt õige arv ja kääne;
3. määrata, millises käändes on lünga ees olev sõna;
4. järjestada käänded;
5. leida küsimusele sobiv vastus;
6. kirjutada õige küsimus ja käände nimetus;
7. moodustada etteantud sõnast sama kääne, kuid mitmuses.

Nii Eesti keele [8] kui ka Sahvri [7] lehel olevates testides on võimalik lahendamise käigus vajutada nuppu „Vihje“, mis annab vihjena ühe tähe, kuid nende kasutamine mõjutab lõppskoori.

Testi kontrollimisel esitatakse punktisumma. Ülesannetes on lausete arv fikseeritud ning korduval lahendamisel on laused samad.

Pärnu Vanalinna Põhikooli eesti keele ja kirjanduse õpetaja Mare Hallop on loonud veebilehekülje [9], kus on kaks käänete õppimise testi [10, 11]. Test „Kääne ja küsimus“ [10] on tehtud veebirakendusega LearningApps [12]. Selles testis tuleb vastavusse panna kääne ja küsimuste grupp. Test „Käänete grupid“ [11] on loodud veebirakendusega Kubbu [13] ning testis tuleb sobitada sõna õige käändegrupiga. Käänete grupid on alguskäänded (nimetav, omastav ja osastav kääne), sisekohakäänded, väliskohakäänded ning viis viimast käänet. Mõlema testi korduval lahendamisel on sõnad ja variandid samad.

1.2 TaskuTark

Veebirakendus TaskuTark [14] sisaldab erinevate ainete videoid, tekste, pildifaile, teste ja loenguid õpilastele 1.-12. klassini. Ka käänete õppimiseks on mitmeid teste, näiteks [14]:

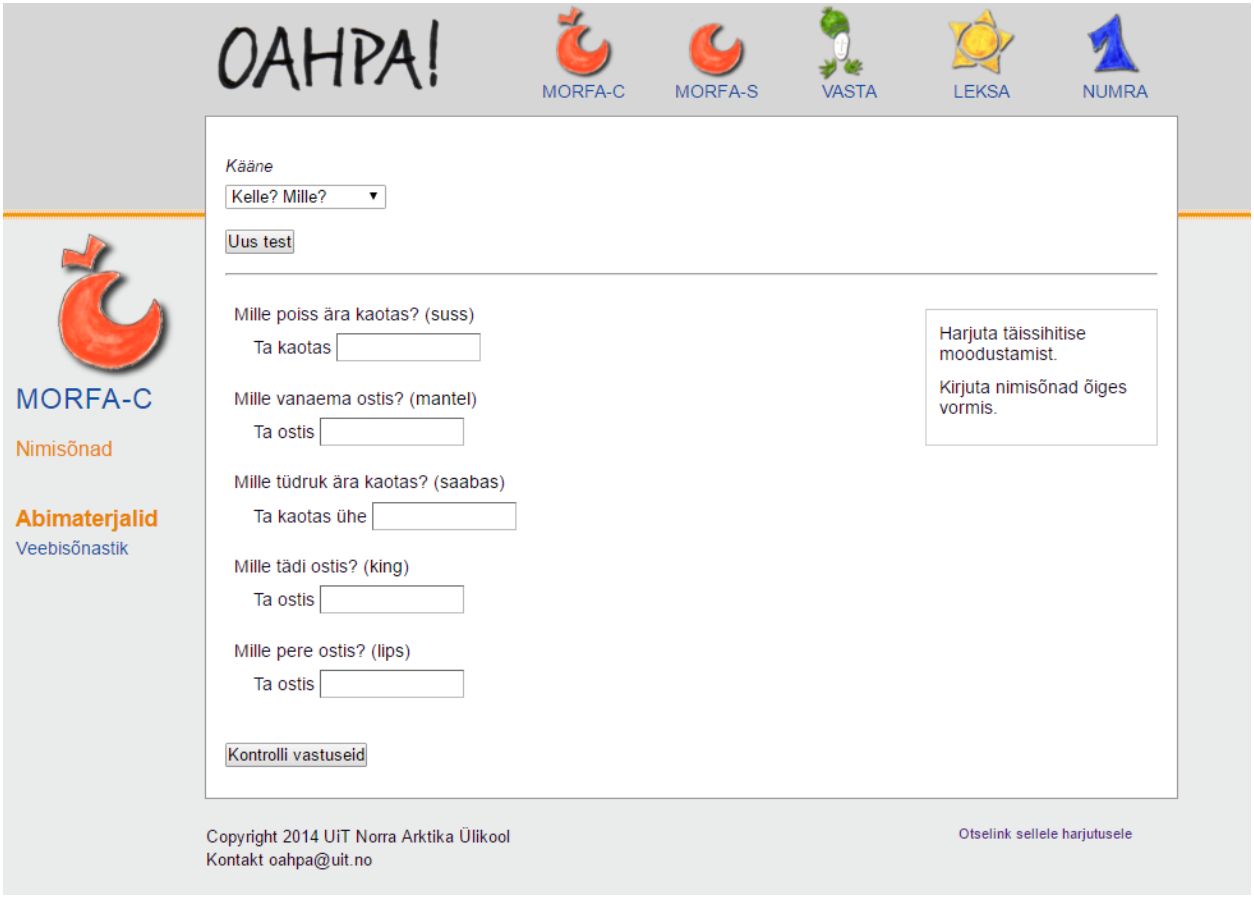
1. „Vali õiged vastused, kuidas käänded jaotuvad“, kus tuleb valida käändeid jaotamise järgi.
2. „Mis käändes ja arvus sõna on?“ (joonis 1), kus on esitatud üks sõna ning kolm valikvastust käände ja arvuga. Sellest lihtsam variant on test „Vali, kas mitmus või ainsus?“.
3. „Vali käändele sobivad küsimused“, kus tuleb valida kõik küsisõnad, mis kuuluvad esitatud käändele. Sama põhimõttega on ka test „Mis küsimused on õiged?“, mis on saava ja nelja viimase käändega.
4. „Vali, mis küsimusele vastab sõna?“, kus tuleb leida sõnale küsisõna vaste.

Joonis 1. Rakenduse TaskuTark test „Mis käändes ja arvus sõna on?“ [14]

TaskuTarga testide ülesanded ei muutu ning pärast ettenähtud testi mahu täitmist teatatakse tulemused ja õiged vastused.

1.3 Oahpa

Õppeprogramm Oahpa [15] on spetsiaalselt keelte õppimiseks loodud rakendus. Oahpa on avatud lähtekoodiga ning loodud paljude erinevate keelte õppimiseks [16]. Eesti Oahpa [15] on arendamisel, sellega on võimalik harjutada käändeid nii sõnade kui ka lausete kontekstis (joonis 2). Joonisel 2 olevas testis on hetkel võimalik valida omastava, osastava, sisseütleva, alaleütleva, seesütleva, alalütleva, saava ja kaasaütleva käände vahel.



Joonis 2. Oahpa lausetega testi vaade [15]

Kõik laused on üsnagi sarnased, sellest tulenevalt on test sobilik algajatele. Sõnadega testis on võimalik valida kolmeteistkümne käände ning 25 peatüki vahel, mis on võetud eesti keele õpikust „E nagu Eesti“. Testides puudub nimetav kääne, kuna nimetavas käändes on sõna ette antud. Samuti on arendamisel võru keele õppimise rakendus [17].

1.4 Grammatika õppeprogrammide võrdlus

Eesti keele käänete õppimiseks leidub nii eelnevalt loodud vahenditega tehtud interaktiivseid teste kui ka spetsiaalselt programmeeritud veebirakendusi. Testide tüüpe on erinevaid. Eeltoodust nähtub, et mitmel analoogilisel õppeprogrammil on olemas test, kus tuleb kääne sobitada küsisõna või küsisõnadega, mistõttu ei ole sarnast testi bakalaureusetöö käigus tehtaval õppeprogrammil tarvilik luua. Sellisel testil on samad vastusevariandid ning uue testi valimise võimalust vaja ei ole.

Ainukese rakendusena on uue testi valimise võimalus rakendusel Oahpa [15], kuid laused ja sõnad hakkavad kiiresti korduma ning need on sarnast tüüpi. Seetõttu on testiga võimalik küll mõnda aega harjutada, kuid pärast paari uue testi genereerimist võib tekkida oht, et kasutajal ei ole enam nii põnev ning harjutamine jääb pooleli. Siinkohal tuleb märkida, et Oahpa on veel demoversioonis.

Kõigil analoogilistel realisatsioonidel tuleb lahendada kogu test ning alles pärast seda näeb tulemust. Kui saada tagasisidet vaid ühe antud vastuse kohta, siis on kasutaja keskendunud ainult sellele ning test täidab paremini harjutamise eesmärki. Samas sobib hindelisele testile paremini korraga kõigi vastuste kontrollimine, sest siis ei ole oluline õppeprotsess, vaid selleks hetkeks saadud teadmiste hindamine.

Lehel Sahver [7] asuvate testide laused on võetud ilukirjandusest. Laused teevad ülesanded elavamaks, lastele lahendamise meeldivamaks ning annab õppimisele juurde lisaväärtuse lugemise näol. Lausetega õppimist pakub ka Oahpa demoversioon [15]. Need laused on lihtsamad ja sobivad hästi lahendamiseks algajale, kes käändeid veel ei tunne.

Torkab silma, et Oahpa [15] ja TaskuTarga [14] ülesanded on intuitiivse kasutajaliidesega ning hea kasutajamugavusega. Lehel Sahver [7] asuvatel raskematel ülesannetel on keeruline aru saada, mida tuleb lünka panna, see võib kasutaja segadusse ajada. Hot Potatoes [6] on aegunud testide loomise vahend, mis ei võimalda luua parema kasutajaliidesega teste.

Üldiselt saab öelda, et eespool kirjeldatud ülesannete ja õppeprogrammide tase on väga erinev. Kindlasti saab neid lahendades esialgse arusaama käänetest, kuid võimalus harjutada pikemalt ja süsteemsemalt puudub ka laiemate võimalustega õppeprogrammidel nagu Oahpa [15] ja TaskuTark [14]. TaskuTargal puudub uue testi genereerimise võimalus. Oahpa demoversioonil ei ole laused veel piisavalt mitmekülgsed.

2. Keeleressursid

Keeleressursid [18] on andmekogumid, mida kasutatakse loomuliku keele uuringuteks ja tehnoloogia arendamiseks. Keeleressursse on erinevaid: tekstikorpused, kõneandmebaasid, leksikaalsed ressursid, tekstitöötlusvahendid ja kõnetöötlusvahendid. Neid kasutatakse keeletehnoloogias, et teostada naturaalse keele analüüsi, sünteesi ja töötlust nii morfoloogilisel, süntaktilisel, semantilisel kui ka pragmaatilisel tasandil [19]. Samuti saab luua dialoogsüsteeme, internetipõhiseid keeleõppeprogramme, avalikele teenustele kasutajaliideseid ja erivajadustega inimestele keeletarkvaralisi abivahendeid [19].

2.1 Tekstikorpused

Tekstikorpused on struktureeritud tekstikogumid, mis on arvutile töödeldavaks tehtud [18] ja neid leiab näiteks Eesti Keeleressursside Keskuse registrist [20]. Järgnevalt on toodud näited eestikeelsetest tekstikorpustest, mis on märgendatud TEI (*Text Encoding Initiative*) põhimõtete järgi [21]:

a) Koondkorpus: Eesti ilukirjandus [22] sisaldab ilukirjanduslikke tekste, kus on ligikaudu 5,8 miljonit sõna ja rohkem kui pool miljonit lauset. Tekstid on avaldatud alates 1990. aastast. Üks fail koosneb ühe teose valitud katkenditest või kogu teosest.
b) Eesti keele segakorpus: Seadused [23] sisaldab Eesti seaduseid, kus on ligikaudu 1,8 miljonit sõna ning Euroopa Liidu õigusaktide eestikeelseid tõlkeid, kus on ligikaudu 9,5 miljonit sõna.
c) Eesti ajakirjanduse korpus [24] sisaldab Eesti ajalehtede artikleid, kus on ligikaudu 182 miljonit sõna.
d) Segakorpus: Riigikogu [25] tekstid on internetist automaatselt salvestatud ning ühes failis on ühe kuu stenogrammid ehk kiirkirjas tehtud üleskirjutused.
e) Segakorpus: Doktoritööd [26] sisaldab doktoritöid ning teadusartikleid.

Tekstikorpuseid on üldiseid kui ka spetsiifilisemaid ning neid saab kasutada erinevatel eesmärkidel.

2.1.1 Korpuste kasutamine keeleõppes

James Wilson on kirjutanud [4], et viimase kahekümne aasta jooksul on kasvanud korpuste hulk, suurused ja kasutatavus. Tema sõnul on lähiaastatel suurenenud korpuste kasutamine ka keeleõppes, kuid vaatamata sellele ei ole see veel piisavalt levinud. Adam Kilgarriff jt on kirjutanud [27], et korpustel on väga suur ning tugev kaudne mõju keeleõppele läbi sõnastike ja õppematerjalide. Ka Wilson tõdeb [4], et korpustest saavad õpetajad võtta testide jaoks lauseid, mis on huvitavad ja autentsed. Samuti väidab Wilson, et korpuste kasutamine keeleõppes ei asenda keeleõppematerjale, vaid tõhusalt täiendab ja toetab neid.

Korpusest keeleõppeks sobilike lausete valimiseks on vaja reegleid, millele lause peab vastama. Seesugused reeglid on loonud Adam Kilgarriff jt [27] konfiguratsioonis GDEX (*Good Dictionary Example*), mis sorteerib sõnastikesse sõnade mõistmiseks ja seletamiseks sobivaid lauseid.

Näiteks kasutasid Adam Kilgarriff jt [27] lihtsamaid parameetreid nagu:

1. lause pikkus on 10 kuni 25 sõna;
2. lause kõik sõnad kuuluvad 17 000 sagedama sõna hulka;
3. lauses ei tohi olla sellised anafoore ja pronoomeneid nagu „this“, „that“, „it“ või „one“;
4. lause peab algama suure tähega ning lõppema puntki, küsimärgi või hüüumärgiga.

Lisaks sellele on Jelena Kallas jt [28] Eesti keele GDEX konfiguratsioonile 1.3 loonud parameetreid, mis aitavad eesti sõnastikesse lauseid sorteerida:

1. lause algab suure tähega ja lõppeb hüüumärgi, küsimärgi või punktiga;
2. lause pikkus peab olema 5 kuni 20 sõna;
3. kui lauses on sees sõna, mille sagedus on alla viie, siis lause sobivus väheneb;
4. lause sõna ei tohi koosneda rohkem kui 20 märgist, vastasel juhul lause sobivus väheneb;
5. lause sobivus väheneb, kui lauses on rohkem kui kaks koma;
6. kui lauses on sees sulud, koolonid, semikoolonid, sidekriipsud, jutumärgid või mõttekriipsud, siis lause sobivus väheneb;
7. lauses peab olema sees tegusõna;
8. kui lauses on ebasobivaid sõnu, siis lause sobivus väheneb;
9. lause sobivus väheneb, kui lause sisaldab pärisnime või lühendeid;
10. lause sobivus väheneb, kui lauses on sõnad „mina“, „sina“, „tema“, „see“, „too“, „siin“, „seal“;
11. lause ei tohi alata sõnadega „mina“, „sina“, „tema“, „see“, „too“, „siin“, „seal“;
12. lause sobivus väheneb, kui lause algab kirjavahemärgiga või sidesõnaga.

2010. aasta seisuga on eesti keel 50 kõrgelt arendatud keeletehnoloogiaga keele hulgas [19] ning eesti keeles leidub korpuseid, mida saab kasutada keeleõppes. Nende kasutamine annab kindlasti eesti keele õppimiseks täiendavat õppematerjali juurde. Eeltoodud parameetreid kohandades on võimalik luua reeglistik, mida saab kasutada ka eesti tekstikorpusest käänete õppimiseks sobivate lausete välja sorteerimiseks.

2.2 Morfoloogiline süntesaator ning analüsaator ja ühestamine

Üheks tekstide töötlemise vahendiks on Pythoni teekide kogumik Estnltk 1.3 [29], millega saab sooritada loomuliku keele töötlemise protseduure, sealhulgas ka morfoloogilist sünteesi ning morfoloogilist analüüsi ja ühestamist.

Morfoloogilise süntesaatoriga genereeritakse etteantud algvormi ja morfoloogiliste kategooriate alusel sõnavorme. Morfoloogilise analüsaatoriga on võimalik määrata ära sõna kääne või pööre, struktuur ja sõnaliik. Tihti võib juhtuda, et tulemuseks tuleb mitu analüüsi varianti, millest kõik ei ole antud kontekstis õiged. Morfoloogilise analüsaatori ühestaja abil valitakse analüüsi tulemustest välja õige variant, arvestades konteksti. Estnltk 1.3 kasutab tõenäosuslikku ühestamist, arvestades kogu sisendiks antud teksti. Seega on vastus täpsem, kui ühestada kogu tekst korraga [29].

2.2.1 Morfoloogilise analüsaatori väljund

Joonisel 3 on Estnltk 1.3 [29] morfoloogilise analüsaatoriga analüüsitud sõna „sepalt“.

```
>>> text = Text('sepalt')
>>> pprint(text.tag_analysis())
{'paragraphs': [{'end': 6, 'start': 0}],
 'sentences': [{'end': 6, 'start': 0}],
 'text': 'sepalt',
 'words': [{'analysis': [{'clitic': '',
                          'ending': 'lt',
                          'form': 'sg abl',
                          'lemma': 'sepp',
                          'partofspeech': 'S',
                          'root': 'sepp',
                          'root_tokens': ['sepp']}],
            'end': 6,
            'start': 0,
            'text': 'sepalt'}]}
```

Joonis 3. Morfoloogilise analüsaatori analüüsi väljund

Käesolevas töös kasutatakse morfoloogilise analüsaatori väljundist järgmist informatsiooni:

1. 'form' – vorm, mille abil saab teada, kas sõna on käänd- või pöördsõna;
2. 'partofspeech' – sõnaliik on vajalik lause struktuurilise ülesehituse kontrollimiseks;
3. 'lemma' – sõna algvorm;
4. 'text' – analüüsitav sõna;
5. 'clitic' – sõna rõhuliide.

Sõna on üheselt määratud, kui morfoloogilise analüsaatori ja ühestaja väljundiks on ainult üks saadud analüüs. Estnltk 1.3 [29] morfoloogilise ühestamise võimalus suurendab oluliselt üheselt määratud sõnade hulka.

2.2.2 Morfoloogilise süntesaatori väljund

Morfoloogilise süntesaatoriga genereeritakse etteantud algvormi ning käände ja sõnaliigi tähise põhjal kõikvõimalikud sõnavormid.

```
>>> synthesize('taim','pl p','S')
['taimi', 'taimesid']
```

Joonis 4. Morfoloogilise süntesaatori väljund

Joonisel 4 on morfoloogiliselt sünteesitud sõna „taim“, et leida mitmuse osastavad vormid, milleks on „taimi“ ja „taimesid“.

2.3 Sagedussõnastik

Sagedussõnastik [30] koosneb 10 000 lemmast, mis on leitud, arvestades sõnade esinemissagedust ühest miljonist ajakirjanduse ja ilukirjanduse sõnast. Ilukirjandusest on võetud allkorpuse tekste mahuga 2 000 sõna, kuid ajakirjandusest kasutatakse kogu ajakirja numbrit. Need kaks tekstiklassi peaksid olema ühtse ja laia levikuga ning neutraalsed. Sõnastikku ei ole võetud ainult kõige sagedasemad sõnad, vaid on arvestatud, et sõna esineks nii ilukirjanduses kui ka ajakirjanduses. Seda arvesse võttes on võimalik tagada, et kõik sagedussõnastikus olevad sõnad on tavalised eesti

keele sõnad, kuid see ei tähenda, et sõnad, mis sagedussõnastikust puuduvad, ei oleks tavalised. Käesolevas töös on vaja sagedussõnastikku, et õppeprogrammi laused koosneksid ainult eesti keeles rohkem kasutatavatest sõnadest.

2.4 Ebasobivate sõnade loend

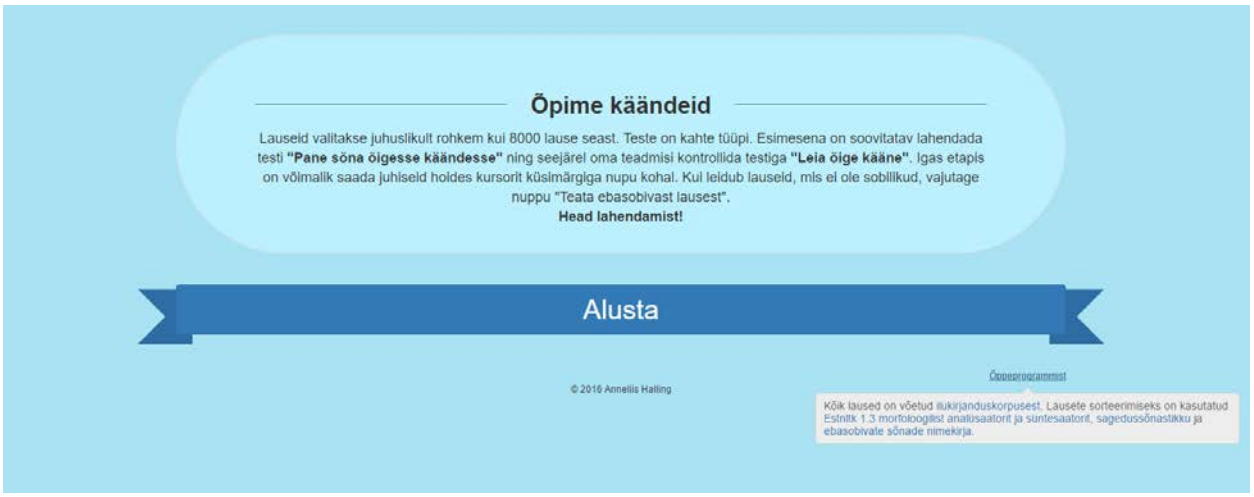
Ebasobivate sõnade loetelu on koostatud Filosofti [31] poolt ning seda on täiendanud Jelena Kallas jt [28]. Lisaks tavalistele eestikeelsetele ebasobivatele sõnadele sisaldab loetelu ka akronüüme ja nii eesti-, inglise- kui ka venekeelseid sõimusõnu.

3. Eesti keele käänete õppeprogramm „Õpime käändeid“

Käesoleva bakalaureusetöö käigus loodud eesti keele käänete õppimise programm[1] põhineb lausetel ning lausetest leitavatel käändsõnadel. Laused, mis on võetud ilukirjanduskorpusest, on autentsed ning nende rohkus võimaldab põhjalikult õppida käändeid. Laused on valitud, kasutades morfoloogilist analüsaatorit koos ühestajaga, sagedussõnastikku, ebasobivate sõnade loendit ning reegleid, mis on loodud spetsiaalselt käänete õppimiseks sobivate lausete sorteerimiseks. Lauseid esitatakse kasutajale juhuslikult rohkem kui kaheksa tuhande lause seast valitud kategooria alusel. Laused annavad õppimisel lisaväärtust lugemise ja sõnavara laiendamise näol. Käändeid on võimalik õppida intuitiivselt testi kaudu, kus lausest üks sõna on asendatud lüngaga. Lisaks on võimalik kontrollida oma teadmisi testiga, kus tuleb määrata, mis käändes ja arvus (ainsus või mitmus) sõna on. Õppeprogramm sobib kõigile vanuseklassidele, kuid mõningad laused võivad olla esimesele ja teisele kooliastmele [32] mõnevõrra keerulised. Antud õppeprogramm on mõeldud süvendatud õppimiseks, seega on keerulised laused kasulikud. Õppeprogrammi kood on kättesaadav GitHubis[2].

3.1 Õppeprogrammi kirjeldus

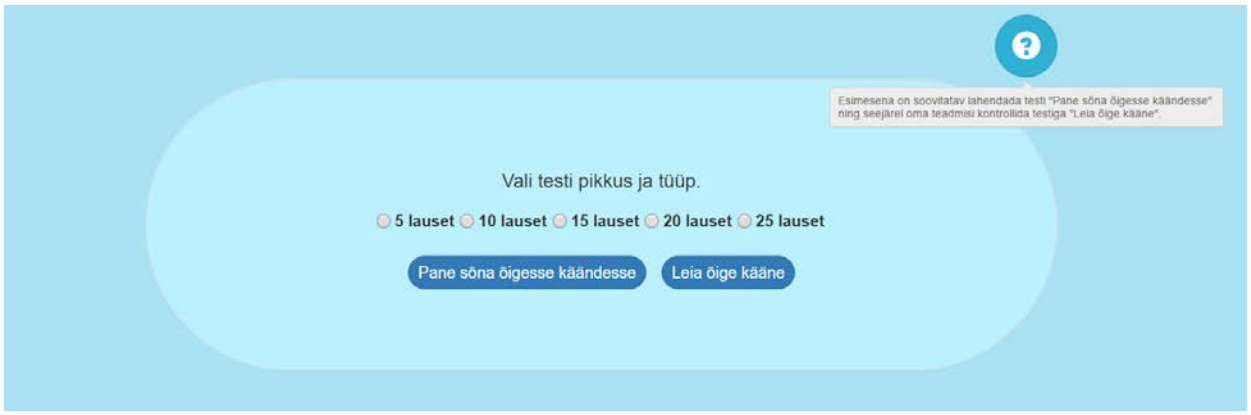
Õppeprogramm koosneb kahest lehest. Joonisel 5 on toodud programmi esileht, kus antakse juhised ja ülevaade.



Joonis 5. Õppeprogrammi esilehe vaade

Vajutades nuppu „Alusta“ kuvatakse kasutajale võimalus valida testi pikkus ja tüüp (joonis 6). Nagu näha joonisel 6, saab kasutaja hiirekursoriga küsimärgile osutades soovituse, millist testitüüpi valida.

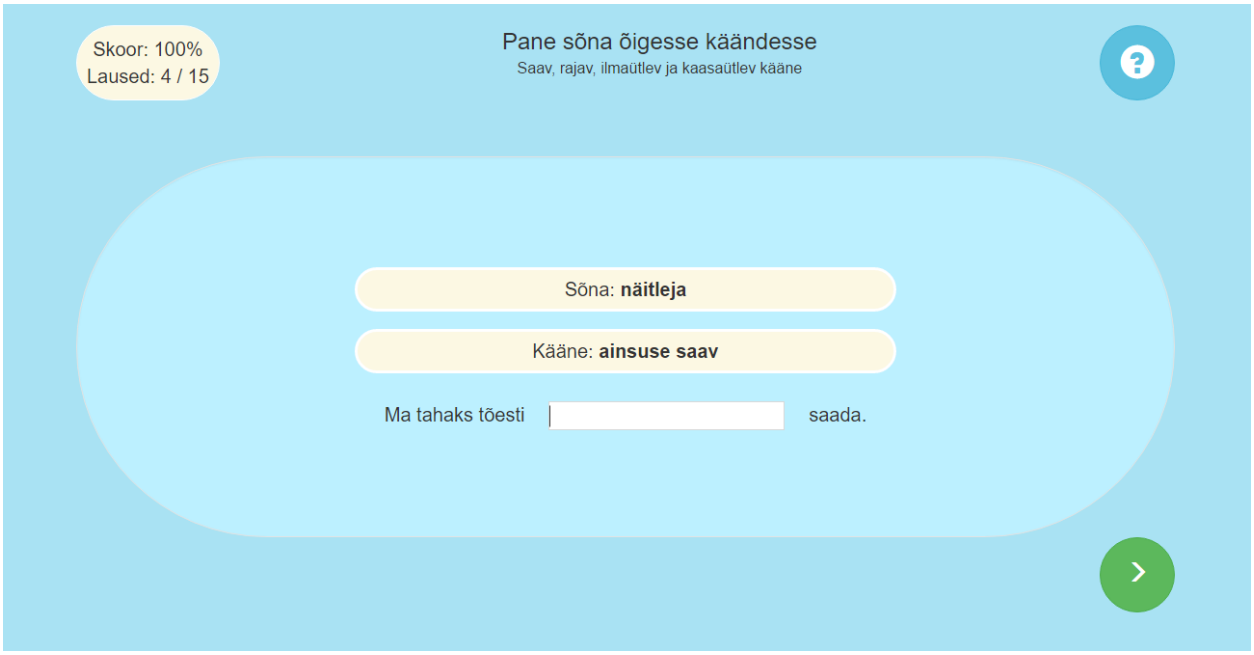
[1] http://prog.keeleressursid.ee/opimekaandeid/
[2] https://github.com/Anneliis/opimekaandeid.git



Joonis 6. Testi pikkuse ja valiku vaade

Õppeprogrammis on kahte tüüpi teste:

1) Testis „Pane sõna õigesse käändesse“ (joonis 7) esitatakse kasutajale lause, kus otsitav käändsõna on asendatud lüngaga. Ette on antud otsitava sõna algvorm ehk sõna nimetavas käändes ning kääne ja arv (ainsus või mitmus), millesse tuleb sõna panna.



Joonis 7. Testi „Pane sõna õigesse käändesse“ vaade

Kui otsitav sõna on koos rõhuliitega, siis on vihjena juures ka rõhuliide -ki või -gi, millega tuleb arvestada sõna käänamisel.

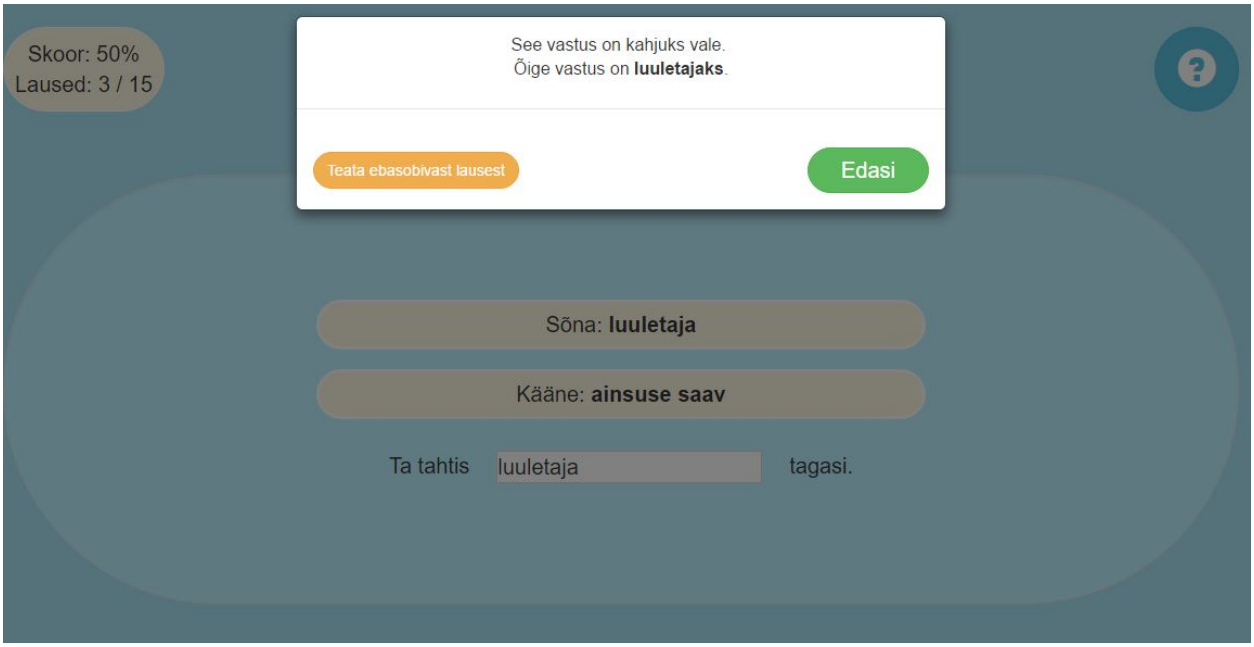
2) Testis „Leia õige kääne“ esitatakse kasutajale lause, kus on üks sõna väljatoodud poolpaksus kirjas (joonis 8). Kasutajal tuleb valida õige kääne neljateistkümne käände hulgast ja määrata, kas sõna on ainsuses või mitmuses.

Joonis 8. Testi „Leia õige kääne“ vaade

Testis „Leia õige kääne“ kasutatakse kõiki käändeid, kuid testis „Pane sõna õigesse käändesse“ on käänded liigitatud vastavalt käänete grammatika õpetamise järjekorrale gruppideks [3]:

- nimetav, omastav ja olev kääne;
- kohakäänded;
- osastav kääne;
- saav, rajav, ilmaütlev ja kaasaütlev kääne;
- kõik käänded.

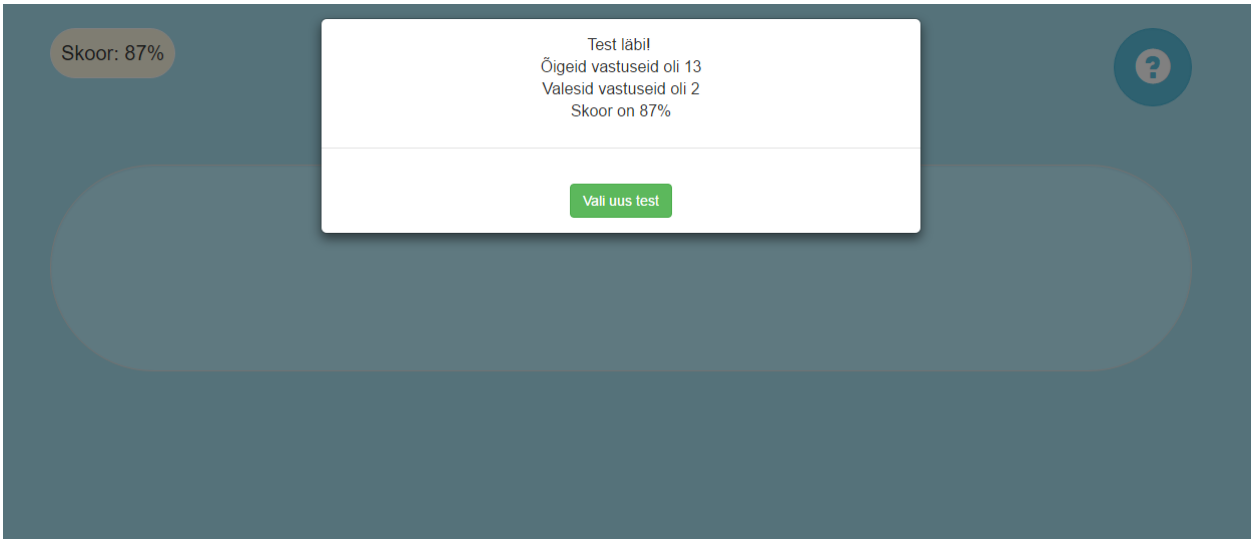
Mõlemas testis on tähtsamate osade väljatoomiseks kasutatud kollast kasti ning poolpaksu teksti, kuna tähtsad detailid peavad olema lihtsasti nähtavad ja selgesti eristatavad [5]. Testi lahendamisel kuvatakse punktisumma protsentides ja näidatakse, mitmes lause on hetkel käsil ning mitu lauset on testis kokku. Samuti kuvatakse testi ja kategooria nimi, mida parajasti lahendatakse. Juhiseid on võimalik saada, osutades kursoriga küsimärgiga sinisele ümmargusele nupule nagu on näha joonisel 8. Pärast vastuse esitamist tuleb kasutajale teade, kas vastus on õige. Kui vastus on õige, antakse sellest kasutajale teada. Kui vastus on vale, öeldakse õige vastus (joonis 9), nii saab kasutaja kohest tagasisidet antud vastusele.



Joonis 9. Vastuse teatamise vaade

Nuppu „Edasi“ ja rohelist noolega nuppu, millega kinnitatakse vastus, on võimalik valida ka sisestusklahviga ning veebilehe laadimisel viiakse tekstikursor alati vastuselahtrisse, mis teeb lahendamise mugavamaks. Testi “Pane sõna õigesse käändesse“ saab lahendada, kasutades ka ainult klaviatuuri.

Korpuses ei ole kõik laused ühtviisi märgendatud, seetõttu võivad mõned laused olla vormi poolest ebasobivad. Samuti leidub lauseid, mis pole sisult kohased. Sellisteks juhtudeks on mõeldud nupp „Teata ebasobivast lausest“.



Joonis 10. Testi tulemuse vaade

Testi lõpus kuvatakse aken (joonis 10), kus on näha, õigete ja valede vastuste arv ning õigete vastuste osakaal. Vajutades nuppu „Vali uus test“ liigutakse tagasi algusesse (joonis 6).

3.2 Õppeprogrammi valitud laused

Laused, mida programm kasutab, peavad olema õpilastele võimalikult arusaadavad, kuid samas piisavalt põnevad, et huvi õppimise vastu püsiks. Eesti ilukirjanduskorpuse [22] laused sobivad õppeprogrammi, kuna korpus on suur ja leidub palju huvitavaid lauseid. Kirjanike sõnavara on tavaliselt lai ning illustreeriv, mis lisaks grammatikale täiendab ka keeleõppija sõnavara. Samuti kasutatakse ilukirjanduslikke tekste eesti keele grammatikat õpetavates õpikutes ja töövihikutes. Seaduseid [23], riigikogu stenogramme [25] ega teaduslikke tekste [26] käesolevas programmis ei kasutata, sest sellised tekstid sisaldavad erialaseid sõnu ega ole keeleõppes üldiselt tarvilikud. Samuti ei sobi ajakirjanduslikud [24] tekstid, kuna ka need laused võivad olla õppimiseks ebasobivad oma keerukuse ja teemavaliku poolest. Eelnevat arvesse võttes on ilukirjanduslikud laused õppeprogrammi jaoks kõige sobivamad.

3.3 Ilukirjanduskorpuse iseärasused

Töö käigus selgus, et ilukirjanduskorpuses [22] ei ole kõik laused ühtmoodi märgendatud. Mõningaid erinevusi on punktuatsioonis ning üldises lausete märgendamises. Leidub juhuseid, kus ühe lausemärgendi sees on mitu lauset, mis on jutumärkides. Näiteks:

1) *<s> “ Madis . ” “ Jah . ” “ Kus sa oled . ” “ Kesklinnas . ” “ Sa tuled kohe siia . ” “ Miks . ” “ Mul on sind ülikiirelt vaja , praegu on hädaolukord . ” </s>*

Samuti on jutumärkide sees olevaid lauseid märgendatud kui eraldi lauseid. Näiteks:

2) *<s> “ Võta jalgadest kinni . </s> <s> Kiiresti . ” </s>*

Järgmises näites puuduvad otsekõnel jutumärgid:

3) *<s> Siis vaatab Endla mulle otsa ning ütleb rõhukalt : Minuga võite kõigest rääkida . </s>*

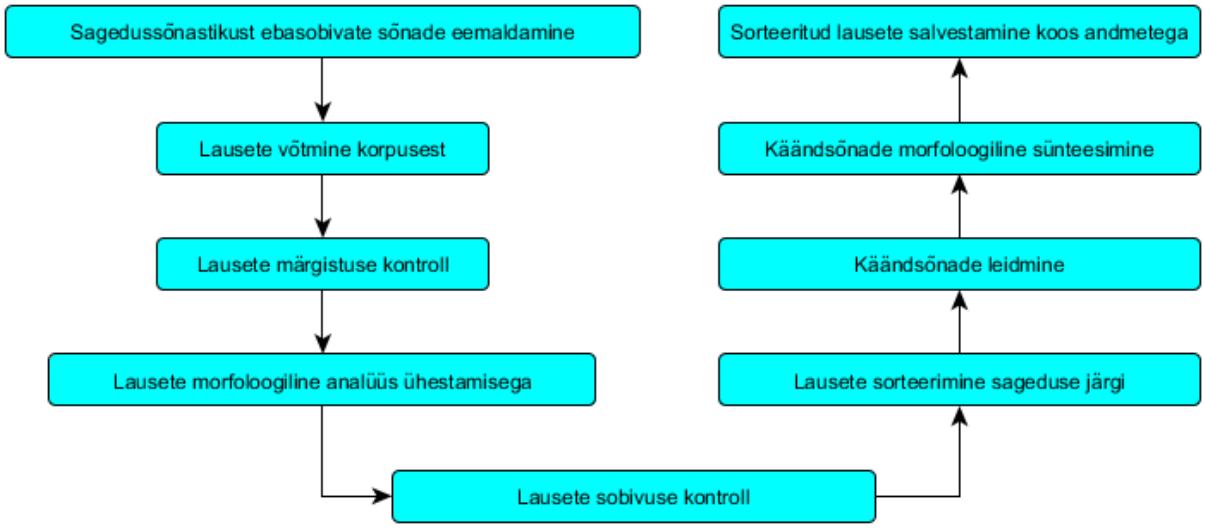
Lause, mille alguses on jutumärk üleliigne:

4) *<s> " " Nälgimine ei võta kaalust maha . " </s>*

Kui võrrelda näidet 2) ja 4), siis selgub, et jutumärgid on erineval kujul. Seesugune varieeruv märgendusviis teeb lausete sorteerimise keerulisemaks.

3.4 Korpuse eeltöötlusprogrammi algoritm

Korpuse eeltöötlusprogramm (joonis 11) on loodud õppeprogrammi jaoks ilukirjanduskorpusest sobivate lausete väljavalimiseks, arvestades ilukirjanduskorpuse eripärasid. Korpuse eeltöötlemine teeb õppeprogrammi töö kiiremaks ja efektiivsemaks. Eeltöötlusprogramm annab kasutajale tehtava töö kohta jooksvalt informatsiooni.



Joonis 11. Eeltöötlusprogrammi algoritm

3.4.1 Sagedussõnastikust ebasobivate sõnade eemaldamine

Sagedussõnastiku kasutamiseks loetakse sisse sagedasemate sõnade loetelu ning eemaldatakse sõnad, mis on ebasobivate sõnade hulgas. Sagedussõnastikus olevaid sõnu kontrollitakse, kuna need võivad olla ebasobivad oma tähenduse poolest. Ebasobivate sõnade loend on täiendatud ka käesoleva töö autori poolt.

3.4.2 Lause võtmine korpusest ja lausemärgistuse kontroll

Korpuse failid on alla laaditud XML TEI P5 kujul [22]. Laused loetakse sisse ühe faili kaupa. Edasi kontrollitakse iga lause kirjavahemärgistust, mis toimub järgmiste sammudena:

1. Esmalt eemaldatakse tühikud lause algusest ja lõpust.
2. Jutumärgid ühtlustatakse, sest korpuses kasutatakse erinevaid jutumärke tähistavaid sümboleid (vt eelpool näide 2 ja 4). Nii on lihtsam jutumärkide esinemist kontrollida ning õppeprogrammis on korrektsem lause esitus.
3. Kui jutumärke on üks või rohkem kui kaks, siis pole lause enam õppeprogrammi sobilik, kuna lause koosneb maksimaalselt seitsmest sõnast, siis ei ole hea, et lause sisaldaks rohkem kui ühte otsekõnet.
4. Lause ei sobi, kui kõrvuti on kaks jutumärki.
5. Lauses ei tohi olla üle kahe koma, sest sel juhul on lause struktuur liiga keeruline.
6. Lause ei tohi sisaldada mõttepunkte, kuna see võib tähendada ära jäetud lause osasid või mõtte pooleli jäämist.

7. Kui terve lause on otsekõne ehk lauses on kaks jutumärki ning ta algab ja lõpeb jutumärkidega, siis need eemaldatakse ning lause sobib õppeprogrammi. Juhul, kui jutumärkide eemaldamisega jäid lause algusesse või lõppu tühikud, siis ka need eemaldatakse.
8. Korpuses leidub faile, kus otsekõne ei ole jutumärkides (näide 4), kuid olemas on otsekõne koolon. Seesuguste lausete vältimiseks eemaldatakse laused, mis sisaldavad koolonit.
9. Korrektsesse 3- kuni 7-sõnalisse keeleõppe lausesse ei sobi ümarsulud, kantsulud, nurksulud, semikoolonid, püstkriipsud, sidekriipsud, kaldkriipsud ja katused. Seega ei võeta õppeprogrammi lauseid, mis sisaldavad selliseid märke.
10. Lause peab lõppema puntki, hüüumärgi või küsimärgiga.

Lauseid kontrollitakse ja muudetakse, kasutades regulaaravaldisi.

3.4.3 Lausete morfoloogiline analüüs koos ühestamisega

Ühe korpuse faili kõikide lausetega, mis läbisid lausemärgistuskontrolli, viiakse läbi morfoloogiline analüüs koos ühestamisega. Laused ühestatakse terve faili kaupa, kuna tõenäosusliku ühestamise vastus on pikema teksti puhul täpsem.

Edasi leitakse morfoloogilisest analüüsist saadud andmed:

1. lauseliikmed;
2. lause sõnad;
3. lause sõnade lemmad;
4. lause sõnade vormid;
5. lause sõnade rõhuliited.

Morfoloogilisest analüüsist saadud andmete abil saab kontrollida lause sobivust.

3.4.4 Lausete sobivuse kontroll

Lausete sorteerimise parameetrid on koostatud, võttes arvesse peatükis 2.1.1 toodud parameetreid, ning muudetud, arvestades käänete õppimiseks sobivaid lauseid:

1. Keeleõppe lause ei pea olema väga pikk, piisab kolmest kuni seitsmest sõnast.
2. Lause peab sisaldama tegusõna.
3. Lauses ei tohi olla kahte samasugust sõna, mis on samas vormis, kuna see annaks kasutajale liiga ilmse vihje.
4. Lause ei tohi alata sidesõnaga, sest sidesõna võib siduda lauset eelneva lausega ning seetõttu pole selline lause ilma kontekstita sobiv.
5. Lause peab algama suure algustähega.
6. Kõik lause sõnad peavad olema üheselt määratud, et sõnadel oleks võimalik määrata üks kindel vorm ja liik.
7. Kõik lauses olevate sõnade lemmad peavad kuuluma sagedussõnastikku, mis määrab ära, et sõna on keeles tihti kasutatav.

Lause sobivuse kontrolli tulemusena jäävad alles ainult võimalikult korrektsed laused, mis sobivad õppeprogrammi.

3.4.5 Lausete sorteerimine sageduse järgi

Korpusest saadud laused sorteeritakse lausestruktuuri järgi sageduse alusel. Erineva pikkusega lausestruktuure arvestatakse eraldi. Lause pikkuse järgi leitakse kõige rohkem esinev

lausestruktuur, mis on selle lause pikkuse puhul maksimumsagedus. Selle järgi arvutatakse miinimumsagedus. Kõik laused, mille struktuuri esinemissagedus on üle miinimumi, võetakse testidesse. Tuleb ära märkida, et lausestruktuuri sagedus avaldab suuremat mõju 3- ja 4-sõnalistele lausetele ning märksa väiksemat mõju 5-, 6- ja 7-sõnalistele lausetele: mida pikem on lause, seda rohkem on erinevaid struktuurivariatsioone.

Laused sorteeritakse lausestruktuuri sageduse alusel, mis näitab kui palju on seda lausestruktuuri ilukirjanduskorpuses. Nii jäetakse välja harva esineva struktuuriga laused. Samuti on võimalik reguleerida sorteeritavate lausete hulka harva esinevate lausete arvelt, langetades või tõstes sageduse lävendit. Töö autor valis miinimumsageduseks 25% maksimumsagedusest, kuna lauseid uurides ning kogust hinnates selgus, et välja jäetakse piisav hulk lauseid, kuid see ei ole kindlasti ainuõige suurus.

3.4.6 Käändsõnade leidmine

Lauses vaadatakse kõik sõnad eraldi läbi ning kontrollitakse peatükis 3.4.3 salvestatud lause andmete kaudu järgmisi aspekte:

1. Lemmade kaudu kontrollitakse, kas sõna on „mina“, „sina“, „tema“, „see“, „too“, „siin“, „seal“. Neid sõnu ei võeta otsitavaks käändsõnaks, kuna lausetes esineb neid asesõnu palju ja seega pole neid mõistlik kasutada.
2. Vormide kaudu kontrollitakse, kas sõna on käändsõna.
3. Juhul, kui sõna on lühikeses sisseütlevas käändes, siis tuleb eraldi ära märkida ainsus, sest lühike sisseütlev kääne on alati ainsuses ning seetõttu puudub sõnavormi noomenkategooriate [29] märgisest ainsus.
4. Kui tegu on nimetava käändega, siis võetakse testi ainult mitmuse vormid, kuna nimetava ainsus on testis „Pane sõna õigesse käändesse“ vihjeks kui sõna algvorm.
5. Mitmuse nimetavas käändes on sõnu väga palju ning seetõttu võetakse iga teine mitmuse nimetava sõnaga lause õppeprogrammi.

Lausest leitud käändsõna, mis vastab eeltoodud nõuetele, sobib õppeprogrammi.

```
<info id="7860">
  <s>Me oleme lihtsad %%% ja täidame käsku.</s>
  <nr>mitmus</nr>
  <case>nimetav</case>
  <n>sõdur</n>
  <word>sõdurid</word>
</info>
<info id="7861">
  <s>Me oleme lihtsad sõdurid ja täidame %%%.</s>
  <nr>ainsus</nr>
  <case>osastav</case>
  <n>käsk</n>
  <word>käsku</word>
</info>
```

Joonis 12. XML-faili kaks kirjet ühe lausega

Kui lauses on mitu sobivat käändsõna, siis lause võetakse õppeprogrammi niimitu korda, kui on sobivaid käändsõnu. Joonisel 12 on näha, et lauses on kaks sõna, mis sobivad otsitavaks käändsõnaks.

3.4.7 Käändsõnade morfoloogiline sünteesimine

Morfoloogilist süntesaatorit on vaja, kuna mõnikord võib ühel käändsõnal olla lühem ja pikem vorm. Seega sünteesitakse leitud käändsõna kõikvõimalikud vormid, et kasutaja sisestatud paralleelvorm loetaks samuti õigeks vastuseks.

3.4.8 Sorteeritud lausete salvestamine koos andmetega ja XML-faili vormindamine

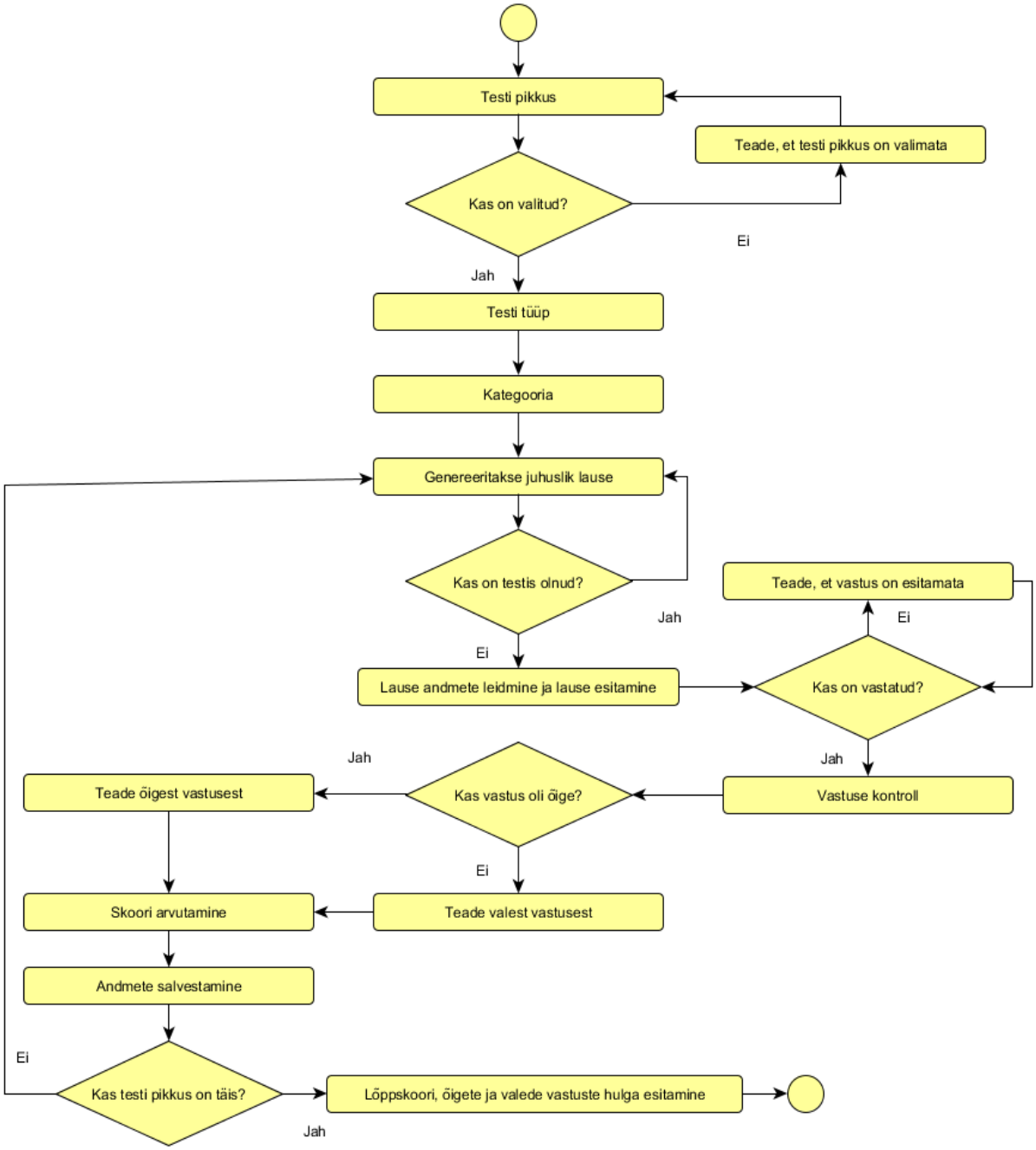
Lause ja leitud käändsõna koos teiste andmetega lisatakse XML-faili, mille märgendid on joonisel 12 ja tähendused on:

1. Kõik andmed on märgendi „info“ sees ja igal infol on oma ID. Kui ühes lauses on mitu sobivat käändsõna, siis salvestatakse laused eraldi unikaalse ID-ga.
2. Sõna, mis algselt oli failis, märgendatakse märgendiga „word“, kuna testis „Leia õige kääne“ esitatakse kasutajale terve lause.
3. Sõnad, mis lisati süntesaatoriga juurde, märgendatakse märgendiga „answer“.
4. Ainsus või mitmus märgendatakse märgendiga „nr“.
5. Lause, kus otsitav või testis olev käändsõna on asendatud „%%%“, on märgendatud märgendiga „s“.
6. Kääne on märgendatud märgendiga „case“.
7. Märgendiga „clitic“ on märgendatud sõna rõhuliide. Kui rõhuliidet ei ole, siis märgend puudub.

Kogu saadud info vormindatakse [33], et XML-failis oleks sisu näha puuna (joonis 12).

3.5 Õppeprogrammi algoritm

Joonisel 13 on õppeprogrammi algoritm, mis kasutab eeltöötlusprogrammiga loodud XML-faile.



Joonis 13. Õppeprogrammi algoritm

3.5.1 Testi pikkus, tüüp ja kategooria

Kui testi tüüp on „Pane sõna õigesse käändesse“, siis tuleb kasutajal valida kategooria, kuid testi „Leia õige kääne“ korral on kategooria „Kõik käänded“. Igal kategoorial on oma XML-fail, sellega

tõuseb kasutamise efektiivsus, kuna failid on väiksemad ning laused on eelnevalt ära jagatud. Test „Leia õige kääne“ ning test „Pane sõna õigesse käändesse“ kategooria „Kõik käänded“ kasutavad sama faili. Lausete hulk, testi tüüp ja kategooria jäetakse meelde ühe testi vältel.

3.5.2 Juhusliku lause valimine ja lause andmete saamine

Programmis loetakse kokku, palju on valitud kategooria järgi failis märgendeid nimega „info“, ning valitakse juhuslik lause. Juhuslike lausete ID-d salvestatakse, et sama ID-ga „info“ märgend ei esineks ühes testis mitu korda. Leitud juhusliku numbriga võetakse XML-failist kõik vajalikud andmed, mis on loetletud peatükis 3.4.8. Kui andmete hulgas on märgend „clitic“, siis esitatakse kasutajale ka info rõhuliite kohta.

Iga kord, kui vastus on sisestatud ning kontrollitud ja kasutaja vajutab nuppu „Edasi“ (joonis 9), genereeritakse uus juhuslik number ja kontrollitakse, ega lause pole hetkel lahendatavas testis varem esinenud.

3.5.3 Vastuse kontrollimine ja lõpptulemuse esitamine

Vastus võib koosneda suur- ja väiketähtedest ning sisaldada alguses ja lõpus tühikuid. Vastavalt valitud testile kontrollitakse vastust. Testis „Pane sõna õigesse käändesse“ loetakse õigeks märgendiga „word“ ja „answer“ märgitud sõnu. Testis „Leia õige kääne“ kontrollitakse, kas valitud on õiged raadionupud. Pärast iga lause kontrolli arvutatakse punktisumma. Testis „Pane sõna õigesse käändesse“ annab iga õige vastus ühe punkti ning testis „Leia õige kääne“ annab kääne pool punkti ja ainsus või mitmus pool punkti.

Kui valitud lausete hulk on täis, kuvatakse teade, et test on läbi, arvutatakse lõppskoor ning õigete ja valede vastuste arv.

3.5.4 Testi andmete salvestamine

Vastuse kontrollimise funktsiooni järel salvestatakse sooritatud katse andmed logifaili. Andmed salvestatakse kausta „data“ alamkaustadesse „word test“ või „case test“. Testi „Pane sõna õigesse käändesse“ logifailid pannakse kausta „word test“ ja andmed kirjutatakse logifaili vastava kuupäevaga. Näiteks 12. mail 2016. aastal on faili nimeks „w-2016-05-12.csv“. Salvestatavad andmed on:

- lause ID;
- lause;
- õige sõna;
- kääne, millesse sõna tuli panna;
- ainsus või mitmus;
- vastus, mis sisestati;
- kas sisestatud vastus oli õige või vale.

Testi „Leia õige kääne“ logifail kirjutatakse kausta „case test“ ning faili nimi on näiteks 12. mail 2016. aastal „c-2016-05-12.csv“. Salvestatavad andmed on:

- lause ID;
- lause;
- küsitud sõna;
- õige kääne;
- õige ainsus või mitmus;

- vastatud kääne;
- kas kääne oli õige või vale;
- vastatud arv (ainsus või mitmus);
- kas arv oli õige või vale.

Iga vastus salvestatakse eraldi, kuna kasutaja ei pruugi alati testi lõpuni teha ning samuti on kiirem salvestada eraldi väiksem hulk andmeid kui kogu test korraga.

3.5.5 Ebasobivate lausete andmete salvestamine

Kui kasutaja teatab ebasobivast lausest, siis lisatakse kausta „data“ CSV-logifail „unsuitableSentences.csv“. Salvestatavad andmed on:

- lause ID;
- lause;
- õige vastus;
- algvorm;
- kääne;
- arv (ainsus või mitmus).

Ebasobivate lausete nimekirja järgi on võimalik käsitsi kustutada lauseid, mis on märgitud ebasobivaks.

3.5.6 Salvestatud andmete kuvamine

Salvestatud andmeid saab vaadata eraldi lehelt[3], kus on olemas testide ja ebasobivate lausete andmed. Andmete tabelis on loodud ka otsinguvõimalus [34]. Kõiki andmeid saab tabeli ülaosast CSV-formaadis alla laadida, mis võimaldab neid tulevikus analüüsida.

Lisaks on samal lehel ka mõlema testi kohta eraldi käänete statistika, kus on võimalik näha testi tüübi järg, kui palju on küsitud erinevates käänetes sõnu, palju õigesti vastatud ning kui suur on õigete vastuste osakaal.

3.6 Tehnoloogilised lahendused

Töö käigus kasutati erinevaid tehnoloogilisi lahendusi nii eeltöötlusprogrammi kui ka õppeprogrammi loomiseks.

3.6.1 Python ja XML

Korpuse eeltöötlust tegev programm loodi programmeerimiskeeles Python, kuna antud keeles on olemas loomuliku keele töötlemise vabavaraline liides [29]. See teeb morfoloogilise analüsaatori kasutamise lihtsaks ja efektiivseks. Korpuse eeltöötlusega on võimalik õppeprogrammi taustal tehtavat tööaega oluliselt lühendada, luues struktureeritud XML-failid, mida kasutab veebiliides.

3.6.2 HTML ja CSS

Veebiliidese kujundamiseks kasutatakse HTML-i ja CSS-i (*Cascading Style Sheets*). HTML-i abil määratakse elementide struktuur ning CSS-i abil muudetakse nende elementide välimust.

[3] Õppeprogrammi "Õpime käändeid" andmed ja statistika http://prog.keeleressursid.ee/opimekaandeid/data.php

3.6.3 Bootstrap

Bootstrap [35] on vabavaraline raamistik, mis sisaldab eelnevalt valmis tehtud CSS-i ja JavaScripti faile ning see lihtsustab veebilehe komponentide välimuse muutmist. Bootstrapi suureks eeliseks on brauseri akna suuruse järgi veebilehe elementide skaleerimine, mis muudab rakenduse kasutatavaks ka mobiilsetel ning väiksema resolutsiooniga seadmetel.

3.6.4 JavaScript, jQuery ja Tooltipster

Õppeprogramm kasutab programmeerimiskeelt JavaScript ning JavaScripti teeki jQuery [36], mis oluliselt lihtsustab JavaScripti kasutamist. JavaScript on populaarne programmeerimiskeel, kõigist internetis olevatest veebilehtedest kasutatakse JavaScripti hinnanguliselt 93,4% [37]. JavaScripti ja jQuery abil on võimalik lugeda XML-faile, kontrollida vastuseid ja lisada või eemaldada HTML-i elemente. Samuti on kasutatud jQuery pistikprogrammi Tooltipster [38], millega on loodud lisainfo kuvamine (joonis 6 ja 8). Tooltipsteris on eelnevalt valmis tehtud JavaScripti funktsioonid. Hoides kursorit komponendi peal, kuvab Tooltipster tekstikasti abiinfoga.

3.6.5 PHP

Serveripoolse programmeerimiskeelega PHP (*Hypertext Preprocessor*) on tehtud andmete salvestamine CSV-faili (*Comma-Separated Values*) ning nende failide kuvamine.

3.7 Õppeprogrammi testimine

Õppeprogrammi testiti 40 korda kahes osas, võttes testi pikkuseks 20 lauset. Kõik testijad täitsid küsimustiku, mille küsimused on näha lisas 1. Esimesel korral testiti 18 korda ning teisel korral 22 korda. Pärast esimest testimist lahendati tekkinud probleemid ning parandati eeltöötlusprogrammi algoritmi. Teisel testimisel olid tulemused oluliselt paranenud.

Kummalgi testimise korral ei tekkinud tehnilisi probleeme. Samuti olid õppeprogrammis olevad juhised kõigile arusaadavad ning kasutajaliides mugav, antud tagasisides sellistele probleemidele ei juhitud.

Esimesel testimisel oli probleeme lausete sobivusega ja vastuste kontrollimisega. Ebasobivast lausest teatati 11 korda ning kokku lahendati 360 lauset. Vastuste kontrollimisel tekkis probleeme, kui otsitav sõna testis „Pane sõna õigesse käändesse“ oli rõhuliitega, kuna õppeprogramm algselt ei arvestanud rõhuliidetega sõnu. Näiteks oli küsitud sõna „päike“ ainsuse alalütlevas käändes, kuid vastust „päiksel“ ei loetud õigeks, kuna nõutud oli vastust „päikselgi“ või „päikeselgi“. Pärast esimest testimist lisati juurde ka rõhuliidete arvestamine. Lisaks sellele oli ebasobivaid lauseid, vaatamata ebasobivate sõnade loetelu abil filtreerimisele, endiselt väga palju. Pärast esimest testimist parandas autor algoritmi ning võttis kasutusele sagedussõnastiku, mis vähendas märgatavalt ebasobivate lausete osakaalu.

Teisel testimisel teatati ebasobivast lausest 4 korda ning kokku lahendati 440 lauset. Nendest kahel korral oli viga selles, et sõna kirjapilt võib olla mitmes käändes ühesugune ja morfoloogilise ühestaja väljund polnud korrektne. Näiteks lauses „Ta taganes kiiresti ja peitis kaamera %%%.“ küsiti sõna „seljakott“ ainsuse osastavas käändes, kuid antud lause kontekstis on sõna „seljakotti“ ainsuse sisseütlevas käändes. Mõlemal käändel on kirjapilt sama. Tegu on morfoloogilise analüsaatori ühestamise veaga. Teised kaks ebasobivat lauset ei olnud head sisu poolest. Ühel lausel jäi mõte kontekstita arusaamatuks, teine lause sisaldas ebasobivat sõna, mis ei olnud ebasobivate sõnade nimekirjas. Selle tulemusena täiendati ebasobivate sõnade loetelu.

3.8 Probleemid

Põhilised probleemid, mis esinesid, olid seotud lausetega ning osad neist ilmnesid ka testimisel. Näiteks võib esineda lauseid, mis on kõikidele parameetritele vaatamata ebasobivad (vt eelpool 3.7). Selleks on loodud nupp „Teata ebasobivast lausest“. Veel ei leita morfoloogilise analüsaatori väljundi ühestamisel alati õiget varianti (vt eelpool 3.7).

Lisaks testimisel ilmnenud probleemidele sünteesib morfoloogiline süntesaator sõna kõik vormid olenemata tähendusest. Näiteks kui testis küsitakse sõna „õlg“ mitmuse osastavas käändes, siis programm loeb õigeks kõik morfoloogilise süntesaatori poolt pakutud variandid olenemata sellest, mis tähenduses on sõna (joonis 14).

```
>>> synthesize('õlg', 'pl p', 'S')
['õlgi', 'õlgu', 'õlgasid', 'õlgesid']
```

Joonis 14. Morfoloogilise süntesaatori väljund sõnaga „õlg“

Kui sisestatakse vales tähenduses sõna, siis on sisestus siiski õiges käändes ning käänete õppimise seisukohalt on kõik korrektne. Samuti on vale tähendusega sõna sisestamise tõenäosus väike.

Eeltoodud probleemid selgusid testimise ning arendamise käigus.

3.9 Edasiarendamise võimalused

Õppeprogrammi testimise tagasiside oli positiivne, kuid kindlasti on olemas edasiarendamise võimalusi, kuna endiselt leidub lauseid, mis ei ole käänete õppimiseks kõige sobivamad. Seega saab edasi arendada eeltöötlusprogrammi algoritmi:

1. Analüüsida ebasobivaks teatatud lauseid ning täiendada lause sobivuse reegleid.
2. Laiendada kasutatavate sõnade valikut, mis teeks lauseid mitmekülgsemaks.

Samuti on võimalik edasi arendada õppeprogrammi:

1. Lisada juurde võimalus valida sõnaliikide vahel, mida käänatakse: nimisõna, omadussõna, arvsõna ja asesõna käänamine.
2. Võrdsustada testis esinevate käänete esinemissagedus.
3. Lisada rohkem mängulisust: kuvada testi sooritamise aeg või piirata seda, luua kasutajakontosid võistlusmomendi tekitamiseks.

Edasiarendamise võimalused, mis vajavad nii eeltöötlusprogrammi kui ka õppeprogrammi edasiarendamist on:

1. Luua raskustasemed, kus on võimalik käänata kahte sõna korraga, näiteks nimisõna koos omadussõnaga.
2. Lisada uusi grammatikateste:
    a. Samalaadselt käänete testidele teha ka pööramise teste, kus oleks sõna antud ma-tegevusnimes ning pööre, millesse sõna tuleb panna.
    b. Lisada rõhuliidete test, kuna rõhuliidete andmed on juba praegu eeltöötlusega loodud XML-failis olemas.

Seesuguseid lahendusi ja võimalusi leidub kindlasti veel, et muuta õppeprogramm kvaliteetsemaks, mitmekülgsemaks ja kaasahaaravamaks.

Kokkuvõte

Käesoleva bakalaureusetöö eesmärgina loodi eesti keele käänete õppeprogramm, mis sisaldab palju autentseid lauseid ning millega saab põhjalikult õppida käändeid.

Töö käigus uuriti, missuguseid keeleressursse kasutada õppeprogrammi loomisel ning leiti, et tekstikorpuste kasutamine on aina populaarsemaks muutuv viis õppematerjalide rikastamiseks. Erinevatest tekstikorpustest valiti välja sobivaim, milleks on ilukirjanduskorpus. Sobivate lausete sorteerimine ilukirjanduskorpusest osutus keeruliseks ülesandeks, kuna lausete sõnaliikide struktuur ei ole alati käänete õppimiseks sobilik ning samuti sisaldavad laused ebasobivaid sõnu. Võimalikult korrektsete lausete sorteerimiseks võeti kasutusele sagedussõnastik ning ebasobivate sõnade loend. Samuti töötati välja reeglid, mis lisaks sagedussõnastikule ja ebasobivate sõnade loendile aitasid leida korrektseid lauseid. Lausete struktuuri kontrollimiseks võeti kasutusele lausestruktuuri sagedus. Keeleressursside ning reeglite kasutamine siiski ei välistanud kõiki ebasobivaid lauseid, mistõttu on testides olemas võimalus teatada ebasobivast lausest. Loodud programmi testiti kahel korral ning pärast mõlemat testimist parandati nii eeltöötlusalgoritmi kui ka õppeprogrammi. Lisaks korpuse eeltöötlusele ning õppeprogrammi veebipõhisele osale lisati juurde ka ebasobivate lausete, salvestatud andmete ja käänete statistika kuvamine.

Kasutatud kirjandus

[1] Entsüklopeedia Eestist. Eesti keel maailma taustal. http://www.estonica.org/et/%C3%9Chiskond/Eesti_keel/Eesti_keel_maailma_taustal/ (11.05.2016).

[2] Eesti keele arengukava 2011−2017. https://www.hm.ee/sites/default/files/eesti_keele_arengukava.pdf (11.05.2016).

[3] Krall, Ingrid; Sõrmus, Elle. 2000. Keeleõpetaja metoodikavihik. Eesti keele grammatika õpetamise võimalusi. Väljaandja: TEA.

[4] Wilson, James. 2013. Technology, pedagogy and promotion. https://www.heacademy.ac.uk/sites/default/files/corpus_technology_pedagogy_promotion2.pdf (11.05.2016).

[5] Digitaalse õppematerjali loomise soovitused. http://oppevara.hitsa.ee/kvaliteet (11.05.2016).

[6] Hot Potatoes. https://hotpot.uvic.ca/ (11.05.2016).

[7] Sahver. https://sahver.wikispaces.com/K%C3%A4%C3%A4nded.+5.+klass.+Testid. (11.05.2016).

[8] Eesti keel. Käänamine. http://eestikeel.eu/kaanamine/ (11.05.2016).

[9] Hallop, Mare. 5. klassi eesti keele töövihik. http://keele6pe5kl.weebly.com/ (11.05.2016).

[10] Hallop, Mare. LearningApps lehel loodud test „Kääne ja küsimus“. http://learningapps.org/display?v=pzamejipc01 (11.05.2016).

[11] Hallop,Mare. E-õppevahendis Kubbu loodud test „Käänete grupid". http://www.kubbu.com/student/?i=1&a=68231_k_nete_grupid (11.05.2016).

[12] LearningApps. http://learningapps.org (11.05.2016).

[13] E-õppevahend Kubbu. http://www.kubbu.com (11.05.2016).

[14] TaskuTark. http://www.taskutark.ee/m/test/ (11.05.2016).

[15] Oahpa eesti keele õppimiseks. http://testing.oahpa.no/eesti/ (11.05.2016).

[16] Oahpa. http://oahpa.no/addlang/index.html (11.05.2016).

[17] Oahpa võru keele õppimiseks. http://testing.oahpa.no/voro/ (11.05.2016).

[18] Eesti Keeleressursside Keskus. https://keeleressursid.ee/et/keeleressursid (11.05.2016).

[19] Riiklik programm. Eesti keeletehnoloogia 2011-2017. https://www.keeletehnoloogia.ee/et/EKT2011-2017-programm-uuendet.pdf/view (11.05.2016).

[20] Eesti Keeleressursside Keskuse register META-SHARE. https://metashare.ut.ee/repository/search/ (11.05.2016).

[21] TEI guidlines. http://www.tei-c.org/Guidelines/ (11.05.2016).

[22] Koondkorpus: Eesti ilukirjandus 1990-. http://www.cl.ut.ee/korpused/segakorpus/eesti_ilukirjandus_1990/index.php?lang=et (11.05.2016).

[23] Eesti keele segakorpus: Seadused. http://www.cl.ut.ee/korpused/segakorpus/seadused/ (11.05.2016).

[24] Eesti ajakirjanduse korpus. https://metashare.ut.ee/repository/browse/eesti-ajakirjanduse-korpus/74b937fc58e311e2a6e4005056b40024c60ddfc6a3054ce196a44cb326d38cf5/ (11.05.2016).

[25] Segakorpus: Riigikogu. http://www.cl.ut.ee/korpused/segakorpus/riigikogu (11.05.2016).

[26] Segakorpus: Doktoritööd. http://www.cl.ut.ee/korpused/segakorpus/doktoritood/ (11.05.2016).

[27] Kilgarriff, Adam; Husák, Milos; McAdam, Katy; Rundell, Michael; Rychlý, Pavel. 2008. Automatically finding good dictionary examples in a corpus. lk 425-432. http://www.euralex.org/elx_proceedings/Euralex2008/026_Euralex_2008_Adam%20Kilgarriff_Milos%20Husak_Katy%20McAdam_Michael%20Rundell_Pavel%20Rychly_GDEX_Automatically%20Finding%20Good%20Di.pdf (11.05.2016).

[28] Kallas, Jelena; Kilgarriff, Adam; Koppel, Kristina; Kudritski, Elgar; Langemets, Margit; Michelfeit, Jan; Tuulik, Maria; Viks, Ülle. 2015. Automatic generation of the Estonian Collocations. https://elex.link/elex2015/proceedings/eLex_2015_01_Kallas+etal.pdf (11.05.2016).

[29] Estnltk 1.3. http://estnltk.github.io/estnltk/1.3/ (11.05.2016).

[30] Sagedussõnastik. http://www.cl.ut.ee/ressursid/sagedused/ (11.05.2016).

[31] Filosoft. http://www.filosoft.ee/ (11.05.2016).

[32] Põhikooli riiklik õppekava. https://www.riigiteataja.ee/akt/114012011001 (11.05.2016).

[33] XML-faili formaatimise funktsioon. https://norwied.wordpress.com/2013/08/27/307/ (11.05.2016).

[34] Andmete sorteerimise JavaScript-i funktsioon. http://jsfiddle.net/giorgitbs/52ak9/1/ (11.05.2016).

[35] Bootstrap. http://getbootstrap.com/ (11.05.2016).

[36] jQuery. https://jquery.com/ (11.05.2016).

[37] W3Techs. Usage of JavaScript for websites. http://w3techs.com/technologies/details/cp-javascript/all/all (11.05.2016).

[38] Tooltipster. http://iamceege.github.io/tooltipster/ (11.05.2016).

Lisad

I. Küsimustik

Õppeprogrammi "Õpime käändeid" testimise küsimustik

Küsimustiku on loonud Anneliis Halling informaatika eriala bakalaureusetöö "Eesti keele keeleressursse kasutav õppeprogramm käänete õppimiseks" testimise raames.

* Kohustuslik

Testimise juhised

Testige testi "Õpime käändeid" Teile määratud testitüübiga ühe korra läbi, valides testi pikkuseks 20 lauset. Kui leidub lauseid, mis on ebasobivad, siis vajutage nuppu "Teata ebasobivast lausest!". Enne testi testimist tuleks lugeda läbi küsimustik, et teada, millele tähelepanu pöörata.
Pärast testimist täitke küsimustik.
Õppeprogramm asub lehel : http://prog.keeleressursid.ee/opimekaandeid
NB! Soovitan proovida ka väljade sisestamata jätmist, et kontrollida õppeprogrammi tehnilist poolt!

Sisestage oma vanus *

Teie vastus

Millist testi lahendasite? *

"Leia õige kääne"

"Pane sõna õigesse käändesse"

Millist kategooriat lahendasite?

Vastake juhul, kui lahendasite testi "Pane sõna õigesse käändesse".

- ○ Kõik käänded
- ○ Osastav kääne
- ○ Omastav ja olev kääne
- ○ Kohakäänded
- ○ Saav, rajav, ilmaütlev ja kaasaütlev kääne

Kas laused olid arusaadavad? *

- ○ Jah
- ○ Ei

Mis jäi arusaamatuks?

Kui vastasite eelmisele küsimusele eitavalt, siis vastake sellele küsimusele.

Teie vastus

Mitu ebasobivat lauset esines? *

Ehk mitu korda vajutasite nuppu "Teata ebasobivast lausest!". Kui ebasobivaid lauseid oli rohkem kui kümme, valige 10.

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Kas testides olevad juhised olid arusaadavad? *

- ○ Jah
- ○ Ei

Mis jäi arusaamatuks?
Kui vastasite eelmisele küsimusele eitavalt, siis vastake sellele küsimusele.

Teie vastus

Kas õppeprogramm kontrollis vastuseid õigesti? *

- Jah, kõik vastused olid korrektselt kontrollitud.
- Ei, leidus mõningaid valesid vastuste kontrolle.

Milliste vastustega esines probleeme?
Kui vastasite eelmisele küsimusele eitavalt, siis vastake sellele küsimusele.

Teie vastus

Kas õppeprogrammis esines tehnilisi vigu? *

- Jah
- Ei

Millised vead esinesid?
Kui vastasite eelmisele küsimusele jaatavalt, siis vastake sellele küsimusele.

Teie vastus

Mis võiks õppeprogrammis veel olla?

Teie vastus

SAADA ÄRA

II. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Anneliis Halling

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose
   Eesti keele keeleressursse kasutav õppeprogramm käänete õppimiseks
   mille juhendaja on Sven Aller,
   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, 12.05.2016