

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Maria Anett Kaha

**EESTI KEELE KEELERESSURSSE KASUTAV
KEELEMÄNG "TUNNE SIHITIST"**

Bakalaureusetöö (9 EAP)

Juhendaja: Sven Aller

Tartu 2024

Eesti keele keeleressursse kasutatav keelemäng “Tunne sihitist”

Lühikokkuvõte:

Bakalaureusetöö eesmärk oli luua eesti keele keeleressursse kasutatav keelemäng “Tunne sihitist”, mis aitab kasutajal õppida või arendada enda grammatilisi oskuseid sihitise käänamises. Teoreetilises osas on lisaks loodud keelemängu protsessi kirjeldusele ka üldine tutvustus olemasolevatest eesti keele keeleressursse kasutatavatest keelemängudest ja nende olulisusest.

Keelemäng on mõeldud kõigile, kes soovivad õppida või arendada enda oskusi sihitise käänamises. Mängu eesmärk on panna ette antud lausetesse õiges käändes sihitised. Pärast mängu valmimist testiti seda ja küsiti kasutajatelt tagasisidet.

Võtmesõnad:

keelemäng, keeleressursside kasutamine

CERCS:

P176 Tehisintellekt, S281 Arvuti õpiprogrammide kasutamise meetoodika ja pedagoogika

Language game “Tunne sihitist” using Estonian language resources

Abstract:

The aim of the bachelor's thesis was to create a language learning game “Tunne sihitist” that helps users learn or improve their grammatical skills in Estonian. The game includes various images related to sentences given to the user for solving. The created language game utilizes existing language resources for the Estonian language. After the completion of the game, tests were conducted, and feedback was collected from users.

In the theoretical part of the bachelor's thesis, in addition to describing the process of creating the game, there is also general information about language games and their importance.

Keywords:

language game, using language Resources

CERCS:

P176 Artificial intelligence, S281 Computer-assisted education

Sisukord

Sissejuhatus	4
1. Lauseliikmed	5
1.1. Sihitis	5
1.1.1. Osasihitis	5
1.1.2. Täisshihitis	6
2. Keelemängud	7
2.1. Keelemängude tähtsus	7
2.2. Keelemängude tüübid	7
3. Keelemäng "Tunne sihitist"	9
3.1. Programmi kirjeldus	9
3.2. Andmed	12
3.3. Eeltöötlus	13
3.4. Piltide eeltöötlus	13
3.4.1. Sobivate piltide leidmine	13
3.4.2. Pildijärjendi genereerimine	14
3.5. Keelekorpuse eeltöötlus	14
3.5.1. Korpuse lausete muutmine tekstiks	15
3.5.2. Klasside järgi jaotamine	15
3.5.3. Sihitist sisaldavate lausete leidmine	16
3.5.4. Lausete korrastamine	16
3.5.5. Klassi failide loomine	17
3.6. Keelemängu "Tunne sihitist" programm	17
3.7. Keelemängu "Tunne sihitist" testimine	19
3.8. Edasiarendamise võimalused	20
Kokkuvõte	22
Viidatud kirjandus	23
Lisad	26
1. YOLO klassid	26
2. Küsimustik	28
3. Litsents	30

Sissejuhatus

Eesti keel on Eesti riigile oluline omariikluse sümbol [1]. Iga keele oskamiseks peab teadma selle keele sõnavara ja lauseehitust. Eesti keel on oma olemuselt keeruline. Põhjuseks võibolla sõnade käänamine, sest eesti keeles on neliteist käänat [2].

Keeleõppes on oluline mõista keelele omaseid lauseliikmeid (alus, öeldis, sihitis, öeldistäide ja määrus), nende järjekorda ja osata neid õigesti käänata [3]. Sihitist tuleb lauses õigesti käänata, kuna see määrab ära, kas lauses kirjeldatud tegevus on lõpetatud või mitte [4].

Tänapäeval on olemas väga palju interaktiivseid õppematerjale ja õpikeskkondi internetis, neist tuntuim on õpikeskkond Opiq [5]. Eesti riikliku õppekava järgi kuulub sihitise õppimine 7.-9. klassi kavasse [6]. Sellises vanuses noored on harjunud olema palju internetis [7]. Interaktiivsed õppematerjalid annavad kasutajale motivatsiooni edasi õppimiseks ja tänapäeva noored eelistavad selliseid õpimeetodeid.

Käesolev bakalaureusetöö on rakenduslik uurimus, mille eesmärk on luua sihitise õppimise jaoks keelemäng, mis kasutab eesti keele keeleressursse. Keelemäng kasutab lisaks tekstikorpusele ka mängu interaktiivseks muutmise eesmärgil pildiandmebaasi. Rakenduslikus uurimuses kasutatakse nii pildituvastust kui ka morfoloogilist analüüsi.

Töö on jagatud kolmeks peatükiks. Esimeses peatükis tutvutakse lauseliikmetega, täpsemalt sihitisega ja tema tüüpidega. Teises peatükis tutvutakse nii keelemängude tähtsusega lapse arengule kui ka keelemängudega, mis kasutavad eesti keele keeleressursse. Viimases peatükis tutvustatakse keelemängu "Tunne sihitist" ja selle jaoks loodud programmi ning kasutatavate andmete eeltöötlust. Lisaks antakse ülevaade testimise tagasisidest ja edasiarendamise võimalustest.

1. Lauseliikmed

Lausetes olevad sõnad jagunevad erinevateks lauseliikmeteks [8]. Lauseliikmed jagunevad pealiikmeteks ja kõrvalliikmeteks. Pealiikmeteks on lauses alus ja öeldis, kõrvalliikmed on sihitis, öeldistäide, määrus ja täiend [3].

1.1. Sihitis

Sihitis on üks lauseliige, mis näitab kellele või millele on lauses olev tegevus suunatud [9]. Tegevuse suunamine tähendab seda, et sihitis on verbist ehk tegusõnast sõltuv nimisõna või nimisõnafraas [3]. Sihitis saab olla lauses omastavas, osastavas või nimetavas käändes [10]. Nende kolme käände järgi jagatakse sihitis osasihitiseks ja täissihitiseks [4]. Osasihitist ja täissihitist mõlemat kasutatakse jaatavas lauses. Eitavas lauses kasutatakse täissihitist ainult siis, kui lause sisaldab “mitte ... vaid”, ülejäänud juhtudel kasutatakse osasihitist. Sihitise valik oleneb tegevuse tulemuslikkusest [4].

1.1.1. Osasihitis

Osasihitiseks (partsiaalobjektiks) nimetakse sihitist, mis on osastavas käändes. Peamiseks eesti keele sihitise käändeks on osastav kääne [6]. Mitmete verbidega saab kasutada ainult osasihitist. Selliseid verbe nimetatakse partitiivverbideks [4]. Partitiivverbideks nimetatakse verbe, mille tegevus ei jõua kindlale tulemusele [4]. Lisaks tavalistele verbidele, näiteks kirjutama, helistama ja sööma, on partitiivverbide seas ka tundeverbid, näiteks kuulma, nägema ja vihkama, ja liigutus-puudutusverbid, näiteks kallistama, lehvitama ja katsuma [4].

Järgnevates näite lausetes on tumedalt välja toodud sihitised. Kasutatud näitelauseid on võetud Eesti keele käsiraamatust (2007) [4].

Laused, mis on õige sihitise käändega:

- Ma meenutasin **minevikku**.
- Leena armastab **lapsi**.

Need samad laused vale sihitise käändega:

- Ma meenutasin **mineviku**.
- Leena armastab **lapsed**.

1.1.2. Täissihitis

Täissihitiseks (totaalobjektiks) nimetatakse sihitist, mis on omastavas või nimetavas käändes [4]. Täissihitis võib esineda ainult sellistes lausetes, kus väljendatud tegevus võib jõuda mingi tulemuseni [4]. See tähendab seda, kui lauses tehtav tegevus on juba teostunud või teostatav ja sihitist viitab kindlale objektile, siis lauses on täissihitis [4]. Verbe, mille puhul kasutatakse täissihitist nimetatakse aspektverbideks [4].

Järgnevates näite lausetes on tumedalt välja toodud sihitised. Kasutatud näitelaused on võetud Eesti keele käsiraamatust (2007) [4].

Laused, milles on kasutatud täissihitist:

- Virve õmbles endale **uue seeliku**.
- Õmble endale **uus seelik**.

Mõlemas näitelauses on täissihitise tingimused täidetud. Esimeses lauses on väljendatud tegevus juba teostatud ehk tegevuseks on õmblemine ja see on juba lõpetatud. Teises lauses väljendatud tegevus teostatakse tulevikus.

Järgnev näitelause on sarnane eelmistega, kuid väljendatud tegevuse tulemuslikkus ei järeldu lausest ehk tegemist on osasihitisega: Virve õmbles endale **uut seelikut** [4].

2. Keelemängud

Parim viis arendada last on mängimise teel, kuna see on üks laste põhitegevusi noores eas. Kristina Tširuhhina [11] kirjutab enda artiklis mängude olulisusest keele omandamisel ja emakeele rikastamisel, rõhutades samal ajal mängude positiivset mõju mõtlemisele, mälule ja peenmotoorikale. Mängimise käigus õpivad lapsed nii uut sõnavara kui ka omandavad erinevaid sotsiaalseid oskuseid. Järgnevalt tutvustatakse uuringuid, mis näitavad mängude olulisust ning tuuakse välja keelemängude olulisemad tüübid.

2.1. Keelemängude tähtsus

Josipa Ključević ja Irena Krumes viisid läbi uuringu [12], kuidas digitaalsete keelemängude kasutamine mõjutab õpilaste motivatsiooni ja tulemusi inglise keele grammatika tundides. Uuringu tulemustest selgus, et õpilased on rohkem motiveeritud õppima inglise keele grammatikat, kui seda tehakse digitaalsete mängude abil. Uuringus osales 114 4. klassi õpilast. Lastele on mängud atraktiivsemad, kui tavaline ülesannete lahendamine õpikust või töövihikust. Seda toetab autorite tehtud statistika. Nimelt hinde “2” saajate osakaal langes 30 protsendilt 10 protsendile ning hinde “3” saajate osakaal langes 42 protsendilt 27 protsendile. Hinde “4” ja “5” saajate osakaal tõusis vastavalt 17 protsendilt kuni 40 protsendini ja 11 protsendilt kuni 23 protsendini [12].

Tänapäeva lapsed on harjunud mängima erinevaid mängu arvutites, telefonides ja konsoolides. Seda suuresti seetõttu, et digitaalsed keelemängud on interaktiivsed ning visuaalsed ja sellepärast on lapsed rohkem motiveeritud õppima just digitaalsete mängudega [12].

2.2. Keelemängude tüübid

Keelemänge on olemas füüsilisel, suulisel ja digitaalsel kujul, pakkudes mitmekülgset ja interaktiivset lähenemist keeleõppele.

Füüsiliste mängude hulka kuuluvad traditsioonilised laua- ja kaardimängud nagu näiteks Alias [13], Scrabble [14] ja Dobble [15]. Sellised mängud ei arenda ainult sõnavara, vaid ka annavad praktilist kogemust keelekasutuses. Näiteks Alias peab seletaja seletama erinevate sõnade või lausete kaudu teistele sõna, mis talle on ette antud.

Suulised keelemängud nagu rollimängud ja sõnamängud, soodustavad suhtlust ja suulise väljendusoskuse arendamist [11]. Suulisi keelemänge kasutatakse väga palju just algkoolides, kus lapsed alles õpivad ennast väljendama. Rollimängudest üks tuntumaid mänge, mida lapsed ja noored mängivad, on Hundimäng [16]. Igal mängijal on kindel roll, mida ta peab täitma. Mängijatest üks on jutustaja, kes juhib kogu mängu. Rollid jagatakse mängijatele kaartide abil laiali. Külaelanikud peavad välja selgitama, kes on libahundid, sest ainult siis on külaelanikud mängu võitnud. Selles mängus on vaja väga palju suhelda, sest see on ainukene viis, et välja selgitada, kes on libahunt. Sellised mängud aitavad peale keelekasutuse arendada ka fantaasiat.

Digitaalsed keelemängud võivad olla efektiivsed vahendid keele õppimisel, pakkudes interaktiivset ja kaasahaaravat õpikogemust. Keelemängud on palju populaarsust kogunud internetis, tuntuim neist on Wordle [17], kus kasutajale antakse iga päev võimalus arvata uut sõna. See paneb inimest tahtma järgmisel päeval uut sõna arvata. Lisaks on ka Eestis loodud palju sõnamänge. Eesti Keeleressursside Keskus [18] pakub enda veebileheküljel erinevaid eesti keele keelemänge, näiteks: Alias Sõnaseletusmäng [19], Õpime käändeid [20] ja Anagrammimasin [21]. Samuti pakub Eesti keele Instituut enda veebileheküljel Sõnaveeb [22] erinevaid keelemänge nii eesti keele sõnade tundmise kui ka sõnade sünonüümide kohta.

Anneliis Hallingu [20] poolt bakalaureuse lõputöö raames loodud interaktiivne käänete õppimise õppeprogramm “Õpime Käändeid”. Sellise õppeprogrammiga on võimalik õppida käändeid nii põhikooli ja gümnaasiumi õpilastel kui ka eesti keelt õppivatel täiskasvanutel.

Õppeprogrammis on võimalik valida kahe harjutuse vahel. Ühel juhul tuleb sõnad õigesse käändesse panna ja teisel juhul valida ette antud sõnale õige kääne. Kasutaja saab ise valida harjutuse mahtu, muutes lausete arvu. Selline õppeprogramm on hea abivahend käänete õppimiseks, sest kokku on 8000 lauset, mille seast kasutajale ülesandeid ette antakse.

3. Keelemäng "Tunne sihitist"

Siin loodud keelemängu "Tunne sihitist"¹ eesmärk on aidata kasutajal õppida sihitise õige vormi lisamist lausesse.

Kasutajale antakse ette juhuslikult rohkem kui 24 000 lause seast neli lauset lahendamiseks ehk ülesanne koosneb neljast lausest. Lisaks lahendatavale lausele kuvatakse kasutajale lauset illustreeriv pilt, millega muudetakse kasutaja jaoks keelemäng interaktiivsemaks. Nendele, kes alles õpivad eesti keelt on hea õppida sõnu, kui nad näevad sellele vastavat pilti. Pildid valitakse juhuslikult 59 pildi seast.

Keelemängule "Tunne sihitist" sobib kõigile, kes soovivad õppida või kontrollida enda teadmiseid sihitise käänamises. Eelkõige on mõeldud keelemäng 7.-9. klassi õpilastele, kuna Eesti riikliku õppekava järgi kuulub sihitise õppimine 7.-9. klassi kavas [6]. Nii keelemängu kui ka eeltöötuse kood koos kasutatavate andmetega on kättesaadav GitHubis²

3.1. Programmi kirjeldus

Keelemängu "Tunne sihitist" veebirakendus koosneb kahest vaatest. Mängu esimeseks vaateks on avaleht. Joonisel 1 on näha avalehe vaade, kus on nupp mängu alustamiseks ja nupp mängu kirjelduse kuvamiseks. Lisaks on märgitud ära kasutatud andmed. Mängu kirjeldus avaneb kasutajale, siis kui kursoriga liigutakse nupu "Mängust" peale, tulemust on näha joonisel 2.



Joonis 1. Avalehe vaade.

¹ <https://kodu.ut.ee/~mariaane/>

² <https://github.com/mariaanett/TunneSihitistM-ng.git>



Joonis 2. Avalehel mängu kirjeldus.

Pärast “Alusta!” nupu vajutamist liigub kasutaja mägu vaatele. Joonisel 3 on näha mängu vaade, kus on lause, millele tuleb valida õige sihitis, sihitise variandid nuppudena ja nelja lauset illustreeriv pilt, millel on lausetega seotud objektid kaetud. Lisaks kuvatakse kasutajale, mitmendat lauset praegu lahendatakse.



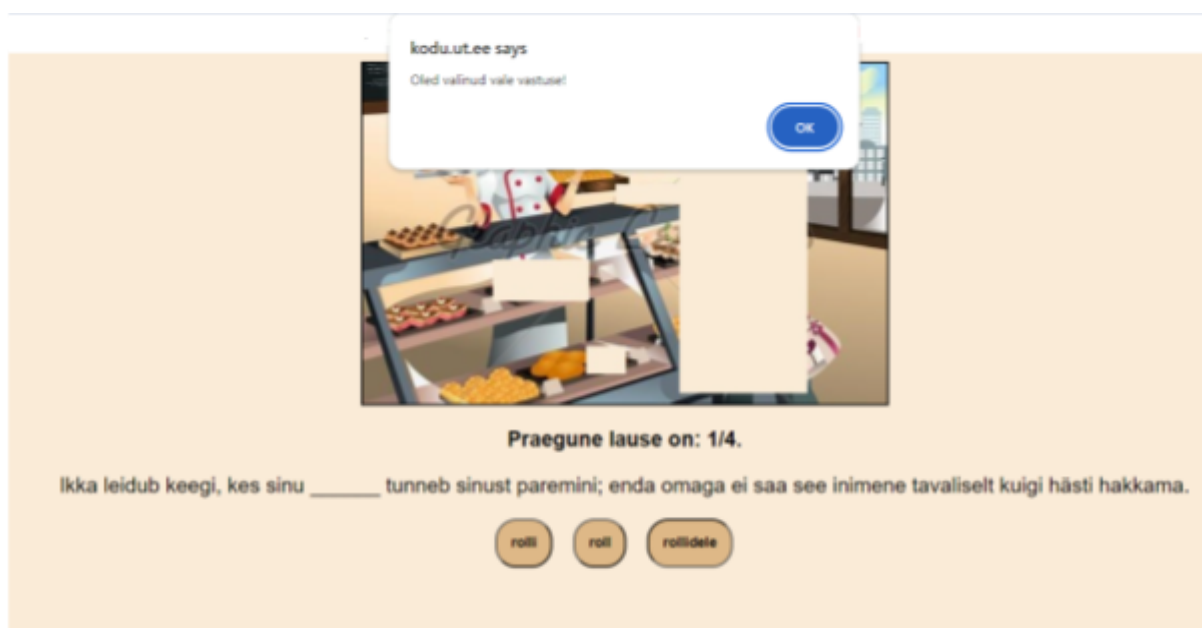
Joonis 3. Mängu vaade.

Peale seda, kui kasutaja on valinud lausesse õige sihitise ehk vajutanud valitud nuppu, siis ilmub ekraanile lause koos sihitisega ning järgmise lause juurde liikumise nupp “Edasi!”, mida on näha joonisel 4. Peale edasi liikumise nupu vajutamist kuvatakse kasutajale uus lause ja pildil avaneb üks peidetud objekt.



Joonis 4. Valiti õige sihitis.

Juhul kui kasutaja valib vale sihitise variandi, siis tuleb ka vastav teavitus, mida on näha joonisel 5. Pärast teavituse ilmumist saab kasutaja selle sulgeda ja uuesti proovida sihitist valida.



Joonis 5. Valiti vale sihitis.



Joonis 6. Mängu lõpp.

Pärast neljanda lause ära lahendamist mäng lõppeb, seda on näha joonisel 6. Kasutajale näidatakse terviklikku pilti. Lisaks loeb programm kokku, mitu valesti vastamist kasutaja tegi, ja näitab vigade arvu. Samuti ilmub nupp “Alusta uuesti”, et kasutajal oleks võimalik uuesti mängida.

3.2. Andmed

Keelemängu jaoks valiti piltide andmebaasiks Clipart1k [23], mis koosneb 1000 eri mõõtmelisest värvilisest pildist. Clipart1k pildid on multifilmi stiilis joonistatud. Pildidel on palju erinevaid objekte ja olukordi. Sellised pildid on sobivad ja lapsesõbralikud antud keelemängu jaoks.

Keelekorpuseks valiti pärast 1990. aastat ilmunud eesti ilukirjanduse korpus [24], mis on eesti ilukirjandustekstide allkorpus, kus on umbes 5,8 miljonit sõna. Kõik selles koondkorpuses olevad allkorpused on XML-faili formaadis. XML (ingl *Extensible Markup Language*) on märgendkeel, mis võimaldab struktureeritud andmete esitamist inimloetaval ja masinloetaval kujul, ning seda kasutatakse laialdaselt erinevate andmete salvestamiseks, edastamiseks ja vahetamiseks [25].

3.3. Eeltöötlus

Selleks, et veebiprogrammi käitusaegne jõudlus oleks parem tehakse kasutatavatele andmetele eeltöötlus ehk andmete puhastamine. Eeltöötluse eesmärk on kasutatavate andmete kvaliteedi parandamine, et need muuta konkreetse ülesande jaoks sobivaks [26]. Boonusena väheneb kasutatavate andmete maht. Lisaks sellele elimineerib kogu eeltöötluse ära tegemine veebiprogrammi käitusaegse Pythoni vajaduse.

Eeltöötluse algoritm tekstikorpusele ja pildifailidele loodi Jupyter Notebookis [27] ning kasutati programmeerimiskeelt Python. Eeltöötluse programmeerimiskeele valik tulenes sellest, et antud keelel on olemas EstNLTK liides, mis pakub eesti keele töötlemise funktsionaalsusi [28]. Jupyter Notebook keskkond valiti sellepärast, et see laseb käivitada koodi osade kaupa, mis on hea eeltöötles, kuna saab lihtsasti tulemusi kontrollida.

3.4. Piltide eeltöötlus

Piltide eeltötlusalgoritm filtreerib Clipart1k pildikaustast välja sobivad pildid. Nii on mängu algoritm palju kiirem, kui jätta välja kõik ebasobivad pildid. Piltide eeltötlusena kõige pealt filtreeriti välja ebasobivad pildid ja siis genereeriti pildile vastav pildijärjend. Pildid valiti kahe piirangu järgi.

Esimeseks piiranguks oli see, et pildil peab olema vähemalt neli erinevat tuvastatud objekti. Objektide arvuks pildil valiti neli, sest pilte, millel on viis või rohkem erinevat objekti, oli liiga vähe. Väiksema arvuga kui neli oleks pilte aga rohkem, kuid siis üks ülesanne koosneks liiga vähestest lausetest, mida kasutaja saaks lahendada iga objekti kohta.

Teiseks piiranguks oli see, et igale pildilt tuvastatud objektile peab vastama vähemalt üks lause, mis seda objekti sisaldab. Peale piltidele tehtud eeltötlust jäi keelemängu "Tunne sihitist" jaoks alles 59 sobivat pilt.

3.4.1. Sobivate piltide leidmine

Sobivate piltide leidmiseks piltide andmestikust oli vaja tuvastada piltidelt nii objektid kui ka objektide arv. Pildilt objektide tuvastamiseks kasutati objektide tuvastajat YOLO (ingl *You Only Look Once*) [30]. YOLO töötab närvivõrkude abil ja tänu oma kiirusele suudab objekte tuvastada ka reaajas. Objekte saab YOLO-ga klassifitseerida 80 erinevasse klassi. Objektide tuvastaja YOLO 80 klassi (vaata lisa 1) tõlgiti käsitsi inglise keelest eesti keelde, et sõnade, mida eeltötlusealgoritm kasutab, tõlge oleks täpsem.

Eeltöötles käib algoritm läbi kõik Clipart1k pildid ja klassifitseerib YOLO objektide tuvastajaga kõik objektid piltide peal. Algoritm valib välja kõik pildid, millel on erinevaid klassifitseeritud objekte vähemalt neli. Sobivad pildid salvestatakse uude kausta.

3.4.2. Pildijärjendi genereerimine

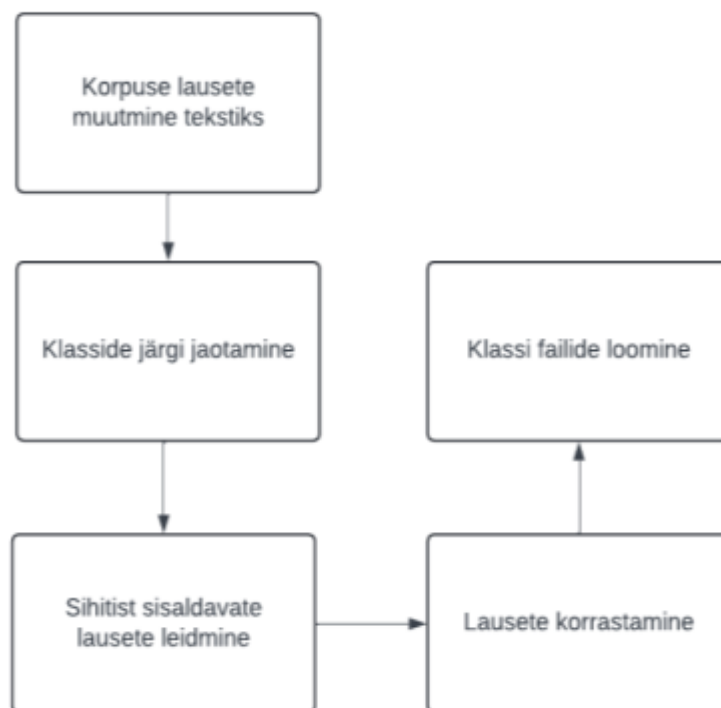
Piltide eeltöötlemise tulemusena saame algse pildi koos tema eri versioonidega, kus on objektid kaetud. See tähendab, et algsest pildist tehakse neli koopiat, kus igal koopial on peidetud vastavalt üks kuni neli tuvastatud objekti ehk esimesel koopial on peidetud üks objekt, teisel kaks objekti jne. Eri versioone algsest pildist on vaja selleks, et peale igat kasutaja poolt vastatud lauset avaneb lausele vastav objekt pildidlt. Nii saadakse vajalik interaktiivne pilt keelemängule.

Algse pildi eri versioonide järjendi genereerimise algoritm töötab järgmiselt:

1. Võetakse eelmises alapeatükis 3.4.1. loodud kaustast pilt.
2. Objektide tuvastaja YOLO-ga otsitakse pildilt koordinaatide järgi üles tuvastatud objektid. Objektid kirjutatakse YOLO klassi nimena (vaata lisa 1) TXT-faili. Loodud tekstifail lisatakse piltidega samasse kausta.
3. Pildilt tuvastatud neli objekti on vaja ära peita. Selleks võetakse algne pilt, millel peidetakse esimene tuvastatud objekt ära, punktis 2 leitud koordinaatide järgi. Saadud pildist tehakse koopia ning salvestatakse faili, seejärel peidetakse ära järgmine objekt, kuni oleme peitnud kõik neli valitud objekti pildilt. Tulemuseks on piltide järjend kus, esimene on algne pilt (kogu pilt on nähtav), teisel pildil on üks objekt peidetud, kolmandal pildil kaks objekti peidetud jne.

3.5. Keelekorpuse eeltöötlus

Keelekorpuse eeltöötlemises oli vaja XML-faili töödelda mängu algoritmile sobivateks TXT-failideks. Nii ei pea iga kord uuesti ebavajalikke, see tähendab YOLO klasse mitte sisaldavaid, lauseid üle käima. Eeltöötlemisalgoritmi on näha joonisel 7. Sobivad laused valiti välja kahe piirangu järgi. Esimeseks piirananguks oli see, et lause peab sisaldama ühte YOLO klassi (vaata lisa 1). Teiseks piirananguks oli see, et lauses peab olema sihitis. Peale lausetele tehtud eeltöötlust jäi keelemängu "Tunne sihitist" jaoks alles 24 098 sobivat lauset.



Joonis 7. Keelekorpuse eeltöötlus

Järgnevalt kirjeldatakse täpsemalt joonisel 7 välja toodud samme.

3.5.1. Korpuse lausete muutmine tekstiks

Ilukirjandus korpuse failidest oli vaja kätte saada ilukirjandus tekstid. Selleks, tuli eemaldada XML-failidest kõik erinevad ebavajalikud päised ja märgendid. Päiste eemaldamiseks kasutati *Beautiful Soup* HTML-parserit [29].

3.5.2. Klasside järgi jaotamine

Sobivate lausete üheks piiranguks oli see, et lause peab sisaldama sõna, mis kattub mõne YOLO (vaata lisa 1) klassi nimega. YOLO klassi sisaldumist lauses kontrolliti morfoloogilise analüüsiga lemmatiseerimise teel, selleks kasutati programmeerimiskeele Python teeki *EstNLTK* 1.7.2 [28]. Alapeatükis 3.5.1. saadud tekst tuli lausete kaupa morfoloogilise analüüsiga lemmatiseerida. Iga lause puhul kontrolliti, kas vähemalt üks sõna lauses kattub YOLO klassiga. Kui selline lause leiti, salvestati lause paisktabelisse, kus võtmeks YOLO klass numbrina (0-79) ja väärtuseks YOLO klassi sisaldavad laused.

3.5.3. Sihitist sisaldavate lausete leidmine

Teiseks sobivate lausete piiranguks oli see, et lause peab sisaldama sihitist. Algoritm käib uuesti laused läbi ja teeb lausete kaupa süntaksi analüüsi. Selleks kasutati EstNLTK 1.7.2 [28] süntaksanalüüsi funktsionaalsust. Kontrolliti, kas lauses esineb selline sõna, mis on sihitise rollis.

Leitud sihitisele genereeriti kaks juhuslikus käändes sihitist. Sihitise kaks varianti sünteesiti kas ainsuses või mitmuses järgnevates käänetes:

- sisseütlev;
- seesütlev;
- seestütlev;
- alaleütlev;
- alalütlev;
- alaltütlev;
- saav;
- rajav;
- olev;
- ilmaütlev;
- kaasaütlev.

Käänete valik tulenes sellest, et sihitis saab ainult olla omastavas, osastavas või nimetavas käändes. Kasutades sihitise variantides nii omastavat kui ka osastavat kääned võib juhtuda, et lausesse sobivad mõlemad käändevormid. Sellisel juhul tekib mängus vastuolu, kus on kaks õiget sihitise varianti.

3.5.4. Lausete korrastamine

Algoritmi poolt leitud sobivad laused olid korrastamata ehk lausete kirjavahe märkidel olid ebavajalikud tühikud. Keelemängu "Tunne sihitist" programmis tuleb kuvada korrastatud lause, sest siis on kasutajal mugavam lauset lugeda.

Näide korrastamata lausest:

Leena ei teadnudki , et neid vanasid ahjusid üldse kütta saab .

Lauses on enne “,” ja “.” tühik, kuid korrektsel lausel nende ees tühikut ei tohi olla.

3.5.5. Klassi failide loomine

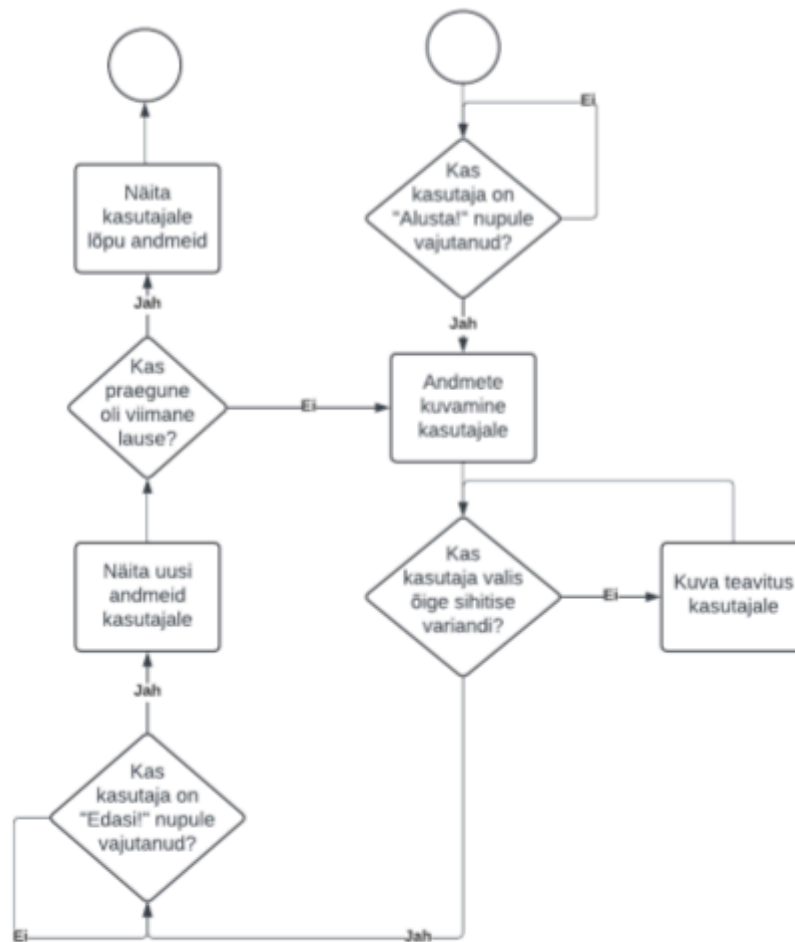
Alapeatükis 3.5.2. loodud paisktabeli abil loodi igale klassile eraldi TXT-fail, kuhu sisse kirjutati kõik vastavad eeltöödeldud laused. Alapeatükis 3.5.3. saadud sihitise variandid kirjutati vastava lause lõppu, eraldatud püstkriipsuga. Sihitise kolm varianti eraldatakse üksteisest semikooloniga. Lause õige sihitise variant kirjutatakse järjekorras esimesena. Näitena on toodud ahi.txt faili lause enne eeltöötlust ja pärast eeltöötlust:

- Leena ei teadnudki , et neid vanasid ahjusid üldse kütta saab .
- Leena ei teadnudki, et neid vanasid ahjusid üldse kütta saab.|ahjusid;ahi;ahjule

Sellises formaadis ridu (iga lause on erineval real) on keelemängu "Tunne sihitist" algoritmil kerge kasutada sellepärast, et vastavas klassi failis on olemas nii laused kui ka nende vastavad sihitise variandid.

3.6. Keelemängu "Tunne sihitist" programm

Järgnevas peatükis kirjeldatakse keelemängule "Tunne sihitist" loodud veebiprogrammi algoritmi. Veebilehe jaoks kasutatakse HTML (*HyperText Markup Language*), JavaScript programmeerimiskeelt ja CSS (*Cascading Style Sheets*). HTML ja CSS kasutatakse veebilehe välimuse kujundamises ning JavaScript-i programmi funktsionaalsuse loomises.



Joonis 8. Keelemängu "Tunne sihitist" algoritm.

Joonisel 8 on kujutatud keelemängu "Tunne sihitist" algoritmi, mis töötab järgmiselt:

1. Kontrollitakse, kas kasutaja on vajutanud nuppu "Alusta!" (vaata joonist 1).
2. Pärast nupule vajutust võetakse juhuslik pilt piltide kaustast. Lisaks pildile võetakse ka tema eri versioonid, kus on objektid peidetud. Seejärel loetakse pildi tekstifailist YOLO klassid, mille järgi otsitakse välja vastavad juhuslikud laused. Lausetest eraldatakse sihitiste variandid ning need pannakse juhuslikku järjekorda. Lisaks eemaldatakse igast lausest sihitis ning asendatakse tühja kohaga.
3. Kasutajale kuvatakse (vaata joonist 3) üks versioon pildist, kus on vastavalt lausele numbrile ära peidetud objektid, ning ka ühele objektile vastav lause. Sihitise variantidest tehakse nupud, et kasutaja saaks valida õige variandi.
4. Kontrollitakse, kas kasutaja on õige sihitise lausesse valinud.
5. Vale sihitise valikul kuvatakse kasutajale vastav teade (vaata joonist 5). Õige valiku puhul kuvatakse kasutajale lause koos õige sihitisega (vaata joonist 4). Kontrollitakse kas kõik laused on kasutajale ette antud, kui on siis kuvatakse kasutajale tema vigade

arv ja nupp “Alusta uuesti”, et saaks uut ülesannet lahendada (vaata joonist 6). Vastasel juhul nupp “Edasi!”, et kasutaja saaks lahendamist jätkata.

6. Peale nupu “Edasi!” vajutamist korratakse samme alates punktist 3.

Algoritm on loodud nii, et kasutaja saaks lõpmatu arv kordi uuesti ülesannet lahendada.

3.7. Keelemängu "Tunne sihitist" testimine

Keelemängu "Tunne sihitist" programmi testimiseks loodi küsimustik (vaata lisa 2). Testimises osales üksteist inimest, kes keelemängu "Tunne sihitist" proovisid ja sellele loodud küsimustikuga tagasisidet andsid. Testijad proovisid mängu vähemalt neli korda, et saada parem ülevaade mängust. Testimise käigus küsiti tagasisidet nii mängu funktsionaalsuse kui ka veebilehe disaini kohta.

Kuna mängu üheks osaks olid illustreerivad pildid, siis oli vaja testida, kas kasutajad panevad tähele, et laused on pildiga seotud. Tagasisidest saadud tulemuste põhjal on pildi seostumist lausega aru saada enamasti siis, kui lause on lihtne. Pikkade lausetega ei pruugi kasutaja aru saada, kuidas lause peidetud objektiga seotud on ehk ei leia lausest sõna, mis on pildil oleva objektiga seotud. Enamus vastajaid aga märkisid, et pildid on heaks interaktiivseks elemendiks mängu juures. Piltide stiili kohta oli positiivne tagasiside, kus mainiti piltide lihtsust, varieeruvust kui ka seda, et pildid olid lõbusad ja värvikirevad. Piltide suurust oleks pidanud fikseerima, kuna tagasisidest saadi, et mõned pildid olid liiga suured ja teised liiga väikesed. Praegusel versioonil on pildidel maksimaalse suuruse piirang. Paar vastanut aga ei pööranud üldse pildile tähelepanu.

Kuna kasutaja eesmärk on leida õiges käändes sihitist, siis oli vaja testida, kas kasutajad saavad aru etteantud lausest ja selle sihitise variantidest. Tagasiside lausete kohta jagunes kaheks osaks: oli olukordi, kus olid enamasti lihtsad ja lühikesed laused, ning oli ka raskeid ja pikki lauseid ette antud. Raskete ja pikkade lausete puhul oli aeganõudvam sihitise kohani jõudmine ja lausest arusaamine, et otsustada, mis peab lünka minema. Selle bakalaureusetöö raames ei piiratud lausete pikkusi, kuna ei osatud oodata, et see kasutajaid segab. Toodi välja ka see, et kuigi laused olid rasked, siis õige sihitise valimine ei olnud keerukas. Testijad arvasid, et laused on mitmekesised, huvitava sisuga ja kirjanduslikud.

Veebilehe disaini poole tagasiside olenes sellest, millisest seadmest testija parasjagu mängu testis. Telefoni kasutajatelt saadud tagasisides oli märgitud, et tekst oli liiga väike ja pidi lugemiseks ekraani suurendama. Kuna mäng loodi arvutikasutajale kasutajasõbralikuks, siis

arvutis testijatel veebilehe disaini poolel probleeme ei esinenud. Üldiselt tagasisidest veebilehe disain on lihtne ja kasutajasõbralik. Mängu juhised olid arusaadavad ja ei tekkinud olukordi, kus kasutaja ei osanud mängus orienteeruda.

Testijate poolt ideid mängule lisamiseks oli erinevaid. Ühe ülesande kohta võiks olla rohkem kui neli lauset, et piltidel oleks rohkem objekte kaetud, kuna hetkel on suurem osa pildist näha. Disaini poolelt nupud võiksid minna vastamise korral vastavalt õigele või valele vastusele kas roheliseks või punaseks.

Testimisel juhtus harva, et kas lause, pilt või sihitise variandid ei ilmunud. Pärast testimist kontrolliti algoritm üle ja leiti viga, mille pärast programm ei tööta vahepeal nii nagu peab. Selgus, et mõnel korral ei jõuta uusi andmeid genereerida, kui juba vaja kuvada neid andmeid. Tehti parandused, et programm töötaks õigesti. Kaks vastajat said ka lahendamiseks nende jaoks mittesobivad laused, mis olid nende arvates rassistlikud ja solvavad.

3.8. Edasiarendamise võimalused

Üldiselt keelemängu "Tunne sihitist" testimise tagasiside oli mitmekülgne ja positiivne, kuid tuleks lisada lausete eeltöötlus algoritmi ebasobivate sõnade list, et ei tekiks olukordi, kus kasutajale antakse ette ebaviisakas või kedagi solvav lause. Testimise käigus avastati ainult kaks sellist lauset, siis võib eeldada, et tekstikorpuse sisaldab veel selliseid lauseid. Sellist sobivate lausete listi on aga keeruline koostada, kuna iga inimene on erinev ja kõigile ei pruugi tunduda samad teemad solvavad või ebaviisakad.

Praegune programm ei arvesta sünonüümidega. Kui YOLO tuvastab pildilt linna, aga lauses on näiteks tuvi, siis see ei ole sobiv. Edasiarendusena saab arvestada selliste sünonüümi (alam- ja ülemmõiste) juhtudega. Selleks saab kasutada näiteks Eesti Wordnet³.

Mängu edasiarenduses on võimalik ka muuta kasutatavaid andmeid. Näiteks saab kasutada teisi keelekorpuseid lausete jaoks, siis saab kasutaja valida, millisest valdkonnast ta lauseid lahendada tahab. Sellisel juhul tuleb arvestada sellega, et YOLO tuvastab kindlaid klasse. Samamoodi võiks valida teistsuguse pildiandmebaasi, kus pildid on samade mõõtmetega ja ei sisalda vesimärke.

Lisaks saab muuta pildi lausega veel rohkem seotuks, kui lauses tuua esile sõna, mis on pildil oleva objektiga seotud, siis tekib ka võimalus nendel, kes alles eesti keelt õpivad, meelde jätta

³ <https://keeleressursid.ee/et/265-eesti-wordnet>

sõnu piltide järgi. Selle jaoks tuleks leitud sõna koos vastava objekti pildiga eraldi pärast vastamist esile tõsta.

Sihitise õppimise edasiarenduseks saab lisada funktsiooni, kus pärast sihitise õigesti valimist seletatakse, miks just selles käändes sinna sobib, ja vastupidi: kui kasutaja valib vale sihitise, siis seletatakse, miks just see sihitis ei sobi lausesse.

Kokkuvõte

Bakalaureusetööna loodi eesti keele keeleressursse kasutav keelemäng "Tunne sihitist", mis aitab kasutajal õppida või arendada enda grammatilisi oskusi sihitise käänamisel.

Töö teoreetilises osas tutvuti keelemängude tähtsusega ja nende kasutamise mõjust laste eneseväljendamisele. Lisaks uuriti, milliseid eesti keele keeleressursse kasutatavaid keelemänge on varem tehtud. Samuti uuriti täpsemalt lauseliikmete ja sihitise olemasolu võimalusi lauses.

Keelemängu "Tunne sihitist" keeleressursiks valiti ilukirjanduskorpus, mille tekstidest filtreeriti välja sobivad laused, kuid laused vajavad veel lisa eeltöötlust ebasobivate sõnade tõttu. Keelemängule lisati interaktiivse osana piltide avamine, mis muutis keelemängu "Tunne sihitist" huvitavamaks. Keelemängule "Tunne sihitist" andsid ka omapoolse tagasiside üksteist testijat. Küsitluse tagasiside oli üldiselt positiivne. Tagasisides välja toodud süsteemsed vead eemaldati.

Viidatud kirjandus

- [1] ERR. Keeleteadlane Anna Verschik: eesti keele olulisus seisneb selle sümbolväärtuses. 2017.
<https://kultuur.err.ee/315988/keeteadlane-anna-verschik-eesti-keele-olulisus-seisneb-sell-e-sumbolvaartuses> (15.05.2024)
- [2] Reili Argus. Ühe minuti loeng: Kas eesti keelt on keeruline õppida? 2015.
<https://novaator.err.ee/256594/uhe-minuti-loeng-kas-eesti-keelt-on-keeruline-oppida> (15.05.2024)
- [3] Mati Erelt, Tiiu Erelt, Kristiina Ross. Lauseliikmed. Eesti keele käsiraamat 2007.
<https://www.eki.ee/books/ekk09/index.php?p=5&p1=1&id=360> (14.04.2024)
- [4] Mati Erelt, Tiiu Erelt, Kristiina Ross. Täis- ja osasihitis. Eesti keele käsiraamat 2007.
https://www.eki.ee/books/ekk09/index.php?link=S%C3%9C_38 (16.04.2024)
- [5] Opiq. <https://www.opiq.ee/> (12.05.2024)
- [6] Vabariigi Valitsus. Põhikooli ja gümnaasiumi riiklik õppekava. Riigi Teataja.
<https://www.riigiteataja.ee/akt/174787> (10.05.2024)
- [7] Eesti Statistikaamet. Noored IT-seadmete ja interneti maailmas.
<https://www.stat.ee/et/uudised/2017/10/26/noored-it-seadmete-ja-interneti-maailmas> (12.05.2024)
- [8] Taskutark. Lauseliikmed.
<https://www.taskutark.ee/study/eesti-keele-tooraamat-8-klassile-i-osa/169696/> (14.04.2024)
- [9] Mati Erelt, Tiiu Erelt, Kristiina Ross. Lauseliikmed. Eesti keele käsiraamat 2007.
<https://www.eki.ee/books/ekk09/index.php?p=5&p1=1&id=360> (14.04.2024)
- [10] Eesti Keele Instituut. Sihitis. <https://eki.ee/teatmik/lause/sihitis/> (16.04.2024)
- [11] Kristina Tširuhhina. Loome, mängime ja õpime eesti keeles. Õpetajate Leht, 2015.
<https://opleht.ee/2015/12/loome-mangime-ja-opime-eesti-keeles/> (03.12.2023)
- [12] Josipa Ključević ; Irena Krumes. Language games as a contemporary method of teaching grammar content in the class context. Horvaatia, Faculty of Education Osijek, 2020.
<https://ikm.mk/ojs/index.php/kij/article/view/728/727> (03.12.2023)
- [13] Alias. About Alias. <https://alias.eu/about-alias/> (01.12.2023)

- [14] Mattel. Scrabble Original. <https://shopping.mattel.com/en-gb/products/scrabble-original-y9596-en-gb> (01.12.2023)
- [15] Dobble. Dobble Classic. <https://www.dobblegame.com/en/games/dobble-classic/> (01.12.2023)
- [16] Brain Games. Hundimäng. <https://www.brain-games.ee/products/hundimang> (16.04.2024)
- [17] Wordle. Wordle Game. https://wordlegame.org/#google_vignette (01.12.2023)
- [18] Eesti Keeleressursside Keskus. Keelemängud. <https://keeleressursid.ee/et/keelemangud> (03.12.2023)
- [19] Sven Aller. Sõnaseletusmäng Alias. <https://keeleressursid.ee/alias/> (14.04.2024)
- [20] Anneliis Halling. Õpime käändeid. <http://prog.keeleressursid.ee/opimekaandeid/> (14.04.2024)
- [21] Eesti Keele Instituut. Anagramm. <https://www.eki.ee/tarkvara/anagramm/> (14.04.2024)
- [22] Eesti Keele Instituut. Sõnaveeb. Keelemängud. <https://sonaveeb.ee/games> (14.04.2024)
- [23] Naoto Inoue. Clipart1k dataset cross-domain-detection, <https://github.com/naoto0804/cross-domain-detection/blob/master/datasets/README.md> (01.12.2023)
- [24] Koondkorpus: Eesti ilukirjandus 1990-. <https://keeleressursid.ee/et/193-eesti-ilukirjandus-1990> (03.12.2023)
- [25] W3Schools. XML Introduction. W3Schools. https://www.w3schools.com/xml/xml_what_is.asp (01.12.2023)
- [26] GeeksforGeeks. Data Preprocessing in Data Mining, <https://www.geeksforgeeks.org/data-preprocessing-in-data-mining/> (16.04.2024)
- [27] Jupyter Notebook. Jupyter Notebook. <https://jupyter.org/> (16.04.2024)
- [28] Estnltk 1.7.2. <https://estnltk.github.io/> (04.03.2024)
- [29] Adomas Sulcas. Beautiful Soup Tutorial - How to Parse Web Data With Python, 2023. <https://oxylabs.io/blog/beautiful-soup-parsing-tutorial> (04.03.2024)

[30] Arxiv. YOLO. <https://arxiv.org/abs/1506.02640> (01.03.2024)

Lisad

1. YOLO klassid

- | | |
|--|--|
| 0: "person", "inimene", | 27: "tie", "lips", |
| 1: "bicycle", "jalgratas", | 28: "suitcase", "kohver", |
| 2: "car", "auto", | 29: "frisbee", "frisbee", |
| 3: "motorcycle", "mootorratas", | 30: "skis", "suusad", |
| 4: "airplane", "lennuk", | 31: "snowboard", "lumelaud", |
| 5: "bus", "buss", | 32: "sports ball", "pall", |
| 6: "train", "rong", | 33: "kite", "tuulelohe", |
| 7: "truck", "veoauto", | 34: "baseball bat", "pesapallikurikas", |
| 8: "boat", "paat", | 35: "baseball glove", "pesapallikinnas", |
| 9: "traffic light", "valgusfoor", | 36: "skateboard", "rula", |
| 10: "fire hydrant", "tuletõrjehüdrant", | 37: "surfboard", "surfilaud", |
| 11: "stop sign", "stop märk", | 38: "tennis racket", "reket", |
| 12: "parking meter", "parkimisautomaat", | 39: "bottle", "pudel", |
| 13: "bench", "pink", | 40: "wine glass", "klaas", |
| 14: "bird", "lind", | 41: "cup", "kruus", |
| 15: "cat", "kass", | 42: "fork", "kahvel", |
| 16: "dog", "koer", | 43: "knife", "nuga", |
| 17: "horse", "hobune", | 44: "spoon", "lusikas", |
| 18: "sheep", "lammas", | 45: "bowl", "kauss", |
| 19: "cow", "lehm", | 46: "banana", "banaan", |
| 20: "elephant", "elevant", | 47: "apple", "õun", |
| 21: "bear", "karu", | 48: "sandwich", "võileib", |
| 22: "zebra", "sebra", | 49: "orange", "apelsin", |
| 23: "giraffe", "kaelkirjak", | 50: "broccoli", "brokoli", |
| 24: "backpack", "seljakott", | 51: "carrot", "porgand", |
| 25: "umbrella", "vihmavari", | 52: "hot dog", "hot dog", |
| 26: "handbag", "käekott", | 53: "pizza", "pitsa", |

54: “donut”, “sõõrik”,
55: “cake”, “tort”,
56: “chair”, “tool”,
57: “couch”, “diivan”,
58: “potted plant”, “potilill”,
59: “bed”, “voodi”,
60: “dining table”, “laud”,
61: “toilet”, “tualett”,
62: “tv”, “televiisor”,
63: “laptop”, “arvuti”,
64: “mouse”, “arvutihiiir”,
65: “remote”, “pult”,
66: “keyboard”, “klaviatuur”,

67: “cell phone”, “mobiiltelefon”,
68: “microwave”, “mikrolaineahi”,
69: “oven”, “ahi”,
70: “toaster”, “röster”,
71: “sink”, “kraanikauss”,
72: “refrigerator”, “külmkapp”,
73: “book”, “raamat”,
74: “clock”, “kell”,
75: “vase”, “vaas”,
76: “scissors”, “käärid”,
77: “teddy bear”, “mänguasi”,
78: “hair drier”, “föön”,
79: “toothbrush”, “hambahari”

2. Küsimustik

Keelemängu Tunne sihitist testimise küsimustik

Küsimustik on loodud Maria Anett Kaha bakalaureusetöö Tunne sihitist keelemängu testimiseks.

Esmalt lugege läbi küsimustik, siis teate, millele testimise ajal tuleks tähelepanu pöörata.

Mängige vähemalt 4 korda läbi.

Keelemäng asub veebilehel: ut.ee/~mariaane

mariakaha@gmail.com [Vaheta kontot](#)



Pole jagatud

* Viitab kohustuslikule küsimusele

Funktsionaalsus

Järgnevad küsimused puudutavad veebilehe funktsionaalsust.

Kirjelda positiivseid ja negatiivseid elemente mängu funktsionaalsuse kohta. *

Teie vastus

Kirjelda lühidalt, kuidas kuvatud pildid illustreerivad lauseid sinu jaoks? *

Teie vastus

Kirjuta positiivseid ja negatiivseid kommentaare piltide kohta. *

Teie vastus

Kirjuta positiivseid ja negatiivseid kommentaare ilmuvate lausete kohta. *

Teie vastus

Kas testimise käigus esines selliseid lauseid, mis ei ole sobilikud keelemängu jaoks? (too ka näiteid) *

Teie vastus

Disain

Järgnevad küsimused puudutavad mängu disaini, kasutajakogemust ja kasutajasõbralikkust.

Kirjelda positiivseid ja negatiivseid elemente mängu disaini kohta. *

Teie vastus

Kirjelda positiivseid ja negatiivseid elemente mängu kasutajasõbralikkuse kohta. *

Teie vastus

Mis võiks mängus veel olla?

Teie vastus

Saada ära

Tühjenda vorm

3. Litsents

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Maria Anett Kaha,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Eesti keele keeleressursse kasutatav keelemäng “Tunne sihitist”**, mille juhendaja on Sven Aller, 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 4.0, 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.

Maria Anett Kaha

15.05.2024