

**KAUNO TECHNOLOGIJOS UNIVERSITETAS**  
**INFORMATIKOS FAKULTETAS**

**Iveta Valentaitė**

**VIRTUALUS ŠNEKAMOSIOS LIETUVIŲ KALBOS PRADMENŲ  
MOKYMASIS PANAUDOJANT BALSO ATPAŽINIMO  
TECHNOLOGIJAS**

Baigiamasis magistro projektas

**Vadovas**  
prof. habil. dr. A. Targamadžė

**KAUNAS, 2016**

**KAUNO TECHNOLOGIJOS UNIVERSITETAS  
INFORMATIKOS FAKULTETAS**

**VIRTUALUS ŠNEKAMOSIOS LIETUVIŲ KALBOS PRADMENŲ  
MOKYMASIS PANAUDOJANT BALSO ATPAŽINIMO  
TECHNOLOGIJAS**

Baigiamasis magistro projektas  
Nuotolinio mokymosi informacinės technologijos (kodas 621E14002)

**Vadovas**

prof. habil. dr. Aleksandras Targamadžė

**Recenzentas**

Doc. Stasys Maciulevičius

**Projektą atliko**

Iveta Valentaitė

**KAUNAS, 2016**



KAUNO TECHNOLOGIJOS UNIVERSITETAS

Informatikos

(Fakultetas)

Iveta Valentaitė

(Studento vardas, pavardė)

Nuotolinio mokymosi informacinės technologijos (621E14002)

(Studijų programos pavadinimas, kodas)

„Baigiamojo projekto pavadinimas“

## AKADEMINIO SAŽINGUMO DEKLARACIJA

20 \_\_\_\_ m. \_\_\_\_ d.

Kaunas

Patvirtinu, kad mano **Ivetos Valentaitės** baigiamasis projektas tema „Virtualus šnekamosios lietuvių kalbos pradmenų mokymasis panaudojant balso atpažinimo technologijas“ yra parašytas visiškai savarankiškai, o visi pateikti duomenys ar tyrimų rezultatai yra teisingi ir gauti sąžiningai. Šiame darbe nei viena dalis nėra plagijuota nuo jokių spausdintinių ar internetinių šaltinių, visos kitų šaltinių tiesioginės ir netiesioginės citatos nurodytos literatūros nuorodose. Įstatymų nenumatytų piniginių sumų už šį darbą niekam nesu mokėjęs.

Aš suprantu, kad išaiškėjus nesąžiningumo faktui, man bus taikomos nuobaudos, remiantis Kauno technologijos universitete galiojančia tvarka.

\_\_\_\_\_  
(vardą ir pavardę įrašyti ranka)

\_\_\_\_\_  
(parašas)

# TURINYS

1. ĮVADAS.....	10
2. ANALITINĖ DALIS .....	12
2.1. Užsienio kalbų mokymosi ypatybės.....	12
2.2. Virtualios lietuvių kalbos mokymosi priemonės.....	13
2.2.1. Priemonių SurfaceLanguages ir Loecsen apžvalga .....	14
2.2.2. Kalbos mokymosi priemonė Duolingo su integruota kalbos atpažinimo funkcija.....	16
2.2.3. Oneness kalbų mokymo kursas .....	18
2.2.4. Lietuvių kalbos mokymosi virtualiu būdu galimybės ir rekomendacijos užsienio studentams .....	18
2.3. Automatinis šnekamosios kalbos atpažinimas .....	19
2.3.1. Paslėptieji Markovo modeliai ir jų pritaikymas kalbai atpažinti .....	23
2.3.2. Garsyno reikšmė automatiniame kalbos atpažinime .....	28
2.3.3. Teksto transkribavimas .....	30
2.4. Analitinės dalies išvados .....	31
3. TYRIMŲ DALIS .....	32
3.1. Tyrimo metodologija.....	32
3.2. Garsynų ir gramatikų paruošimas pavardžių ir vardų atpažinimo tyrimams .....	33
3.3. Pavardžių ir vardų apmokyto profilio „Default_Iveta1” įtaka atpažinimo tikslumui .....	35
3.4. HTK atpažinimo programų rinkinio praktinis pritaikymas ir įsisavinimas.....	37
3.4.1. Pavardžių ir vardų tyrimų su HTK rezultatų analizė.....	44
3.5. Tyrimų dalies išvados.....	47
4. PROJEKTINĖ DALIS .....	48
4.1. Lietuvos studentų poreikių analizė.....	48
4.2. Dėstytojų apklausos analizė .....	50
4.3. Užsienio studentų poreikių analizė.....	52
4.4. Lietuvių kalbos mokymo kurso metodika .....	53
4.5. Sistemos struktūra .....	54
4.6. Sistemos realizacija .....	56
4.7. Sistemos kokybės tyrimas .....	57
4.8. Projektinės dalies išvados.....	59
IŠVADOS .....	60
LITERATŪRA.....	61
1 PRIEDAS Lietuviškų pavardžių ir vardų garsynas.....	64

2 PRIEDAS Pavardžių ir vardų atpažinimo tikslumo priklausomybė nuo būsenų skaičiaus .....	65
3 PRIEDAS Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus ir +4 būsenų.....	66
4 PRIEDAS Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus ir fiksuoto 14 būsenų skaičiaus.....	67
5 PRIEDAS Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus ir fiksuoto 16 būsenų skaičiaus.....	68
6 PRIEDAS Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus ir fiksuoto 17 ir 21 būsenų skaičiaus.....	69
7 PRIEDAS Netaisyklingai išstartų pavardžių ir vardų atpažinimo tikslumo priklausomybė nuo Gauso skirstinių ir fiksuoto būsenų skaičiaus .....	70
8 PRIEDAS Balso komandos gramatikai nr. 1 .....	71
9 PRIEDAS Gramatika nr. 1 .....	72
10 PRIEDAS Balso komandos gramatikai nr.2 .....	76
11 PRIEDAS Gramatika nr. 2.....	77
12 PRIEDAS Balso komandos gramatikai nr.3 .....	80
13 PRIEDAS Gramatika nr. 3.....	81
14 PRIEDAS Lietuvos studentų apklausos klausimai .....	84
15 PRIEDAS Lietuvos studentų poreikių analizė (pilna) .....	85
16 PRIEDAS Dėstytojų apklausos klausimai .....	90
17 PRIEDAS Dėstytojų apklausos analizė (pilna).....	91
18 PRIEDAS Užsienio studentų apklausos klausimai .....	96
19 PRIEDAS Užsienio studentų poreikių analizė (pilna).....	97
20 PRIEDAS Kokybės tyrimo apklausos klausimai.....	103
21 PRIEDAS Kokybės tyrimo analizė (pilna) .....	104

Valentaitė, Iveta. Virtualus šnekamosios lietuvių kalbos pradmenų mokymasis panaudojant balso atpažinimo technologijas. Magistro baigiamasis projektas / vadovas prof. habil. dr. Aleksandras Targamadžė; Kauno technologijos universitetas, Informatikos fakultetas.

Mokslo kryptis ir sritis: Technologijos mokslai, informatikos inžinerija (07T)  
Reikšminiai žodžiai: *kalbos atpažinimas, virtualus kursas, garsynas, HTK, ispanų kalbos atpažintuvas, metodika.*

Kaunas, 2016. 104 p.

## SANTRAUKA

Lietuvos universitetus nuolat gausina užsienio šalių studentai, kalbos mokėjimas jiems tampa neatsiejamas nuo kasdienio gyvenimo. Kalbos išmokimas svečioje šalyje gali tapti nesudėtingu mokymosi procesu, mokantis virtualiu būdu. Mokymosi procese dalyvauja tik besimokantysis jam patogiu laiku.

Šiai problemai spręsti pasitelkiama virtuali mokymosi aplinka, kurioje, remiantis sukurta mokymosi metodika, talpinama mokymosi medžiaga ir pritaikomos informacinių technologijų priemonės, siekiant užtikrinti lietuvių kalbos: leksikos, gramatikos, fonetikos ypatybių įsisavinimą. Siekiant išsiaiškinti lūkesčius kuriamam kalbos mokymosi kursui, virtualiame kurse teiktinų metodų svarbą ir efektyvumą, atliktos ekspertų ir studentų apklausos. Taip pat analizuotos internetinės kalbos mokymo priemonės, jų naudojamas kalbos mokymo principas, pateikiamų užduočių tipų įvairovė. Tai daro įtaką parenkant kalbos mokymo medžiagos turinį, išdėstymą, mokymo metodus ir IT priemones.

Sukurta sudėtingai užsieniečiams tariamų pavardžių ir vardų garsynas, kurį sudaro daug diakritinių ženklų turinčios raidės. Atlikti garsynų tyrimai, taikant įvairius parametrus ir sąlygas. Tyrimai atlikti su HTK ir ispanų kalbos atpažintuvu, siekiant išsiaiškinti sudėtingų balso komandų atpažinimo tikslumą. Tai naudojama nustatant, kokios sudėtingos frazės atpažįstamos taisyklingai, kad vėliau tuo remiantis būtų galima sukurti gramatikos failus su balso komandomis pagal mokymosi kursą, kurį naudos ispanų kalbos atpažintuvu besiremianti sukurta programa. Programa ekrane atvaizduos atpažintą balso komandą ir atpažinimo matą.

Virtualiame kalbos mokymo kurse kalbėjimo įgūdžiams lavinti taip pat pritaikoma *speech to text* (kalba verčiama tekstu) principu besiremianti programa, naudojanti balso įrašymo ir atkūrimo funkciją. Teoriniams kalbos dalykams aiškinti remiamasi kalbos mokymosi projekto „Oneness“ projekto medžiaga, talpinami kryžiažodžiai, sukurti „HotPotatoes“ įrankiu, naudojami vaizdo įrašai.

Valentaitė, Iveta. *The Virtual Learning Of The Elements Of The Spoken Lithuanian Language By Means Of Voice Recognition Technology*: Master's thesis in Information Technologies of Distance Education / supervisor assoc. prof. Aleksandras Targamadzė. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Technological Sciences, Informatics Engineering (07T)

Key words: Speech recognition, virtual course, speech corpora, methodology, HTK toolkit, Spanish recognition toolkit

Kaunas, 2016. 104 p.

## SUMMARY

Lithuania attracts more and more foreign students so that language knowledge becomes the essential part of their daily life. Language acquisition in foreign country can become simple learning process when students are learning virtually at a suitable time for them.

To solve this issue virtual learning environment is chosen according created learning methodology, containing learning material and applied information technology tools to ensure assimilation of the features of Lithuanian lexis, grammar and phonetics. In order to ascertain expectation for developed language learning course the surveys of experts and students have been done to find out the importance and efficiency of methods of virtual course. In addition, internet-based language teaching tools, their used language teaching methods, the variety of presented tasks have been analyzed too. This affects the choice of content of teaching material, arrangement, methods and IT tools.

The speech corpora of difficult surnames and names having a lot of diacritical marks and other linguistic features has been analyzed applying diverse parameters and conditions that have been created. The analysis has been done with HTK and Spanish language recognition toolkit to find out the accuracy of difficult voice recognition commands. This is used for determining what difficult phrases are recognized accurately and later on this basis, grammar files with voice commands according learning course can be created and used by Spanish language recognition program. The program shows the dictated phrases or recognition accuracy on the screen.

In the virtual language teaching course to develop speaking skills the program according to *speech to text* principle which uses voice recognition and re-creation function is applied. To clarify the theoretical material of language the material of "Oneness" language learning project is based on, crosswords created by "HotPotatoes" tool are placed and videos that are used.

## **LENTELIŲ SĄRAŠAS**

1 lentelė. Tyrimo su teisingomis balso komandomis rezultatai ..... 36

2 lentelė. Tyrimo su iškraiptomis balso komandomis rezultatai..... 36



## PAVEIKSLŲ SĄRAŠAS

<b>1 pav.</b>	Automatinio kalbos atpažinimo sistemos struktūra .....	20
<b>2 pav.</b>	Tipinės paslėptųjų Markovo grandinių topologijos .....	24
<b>3 pav.</b>	HTK atpažinimo įrankio šnekos apdorojimo schema.....	26
<b>4 pav.</b>	Pavardės ir vardo „adomavičiūtė inga” modelių failas .....	39
<b>5 pav.</b>	Balso komandos „adomavičiūtė inga” modelių failas (+4 būsenos ir 6 skirstiniai).....	42
<b>6 pav.</b>	Efektyviausi kalbų mokymosi metodai.....	49
<b>7 pav.</b>	Žinios mokymosi procese .....	49
<b>8 pav.</b>	Įgūdžiai kalbos mokymosi procese.....	51
<b>9 pav.</b>	Efektyvių kalbos užduočių tipai .....	51
<b>10 pav.</b>	Pageidautina išmokti kalba .....	52
<b>11 pav.</b>	Virtuali lietuvių kalbos mokymo sistema .....	55
<b>12 pav.</b>	Mokymosi medžiagai patrauklumą teikiantys aspektais .....	58

## 1. ĮVADAS

Kasmet auga į Lietuvos universitetus atvykstančių užsienio studentų skaičius. Pastebima, kad pastaraisiais metais smarkiai kito šalių, iš kurių studentai atvyksta į Lietuvą, sąrašas. Anksčiau studijas Lietuvoje dažniausiai buvo linkę rinktis studentai iš Azijos žemyno, tačiau daugėja iš Europos šalių atvykstančių studentų. Svetimos kalbos nemokėjimas ir šalies nepažinojimas užsienio piliečiams gali sukelti sunkumų bandant prisitaikyti naujoje aplinkoje. Nauja aplinka ir kalba užsienio piliečiui gali atrodyti kaip natūralus dalykas (ypač tam, kuris yra linkęs greitai prisitaikyti prie naujų sąlygų), tačiau nauji šalies komponentai (kalba, kultūra ir pan.) gali sukelti nepatogių situacijų ir problemų. Taip pat svetimos tautos nesuvokimas gali pasireikšti kultūriniu šoku. Svarbiausia svetimose šalyse – bent kiek išmanyti kalbą, nes tai leidžia lengviau pritapti prie visuomenės. Nemokantis lietuviškai skaityti žmogus, negali perskaityti įvairių užrašų miesto informacinėse lentose, informacijos universitete, jaučiasi tarsi atskirtas nuo visuomenės, nes pramogos (filmai, televizija, radijas ir pan.) taip pat neprieinamos. Galima išvelgti ribotas galimybes lankantis teatruose, parodose ar koncertuose, o, svarbiausia, kiekvieną dieną lankomose vietose – parduotuvėse.

Užsienio studentų galimybių apribojimą lemia ir tai, kad nėra galimybės gauti informacijos iš praeivių, nes į jiems išskylančius klausimus gali atsakyti žmonės, mokantys šiems studentams suprantamą kalbą – dažniausiai anglų kalbą. Tačiau Lietuvoje nedaug vyresnio amžiaus žmonių moka angliškai, o ir ne visas jaunimas kalba angliškai. Atsižvelgiant į šią ir kitas panašias situacijas pravartu bent kiek mokėti kalbą tos šalies, į kurią važiuojama mokymosi ar kitais tikslais.

Svetimos kalbos išmokimas šiandieninėje visuomenėje – tai neatskiriama nuo šiuolaikiško, laisvo ir ryžtingo žmogaus. Tai proga gyventi globaliai, daugiakultūriu lygmeniu bei turėti plačių galimybių. Kalbos išmokimas XXI-ame technologijų amžiuje tampa įdomiu bei visiems prieinamu dalyku, nes minėtojo proceso metu yra naudojamos įvairios technologijos, kurios leidžia įsisavinti informaciją dar prieš vykstant į svetimą užsienio šalį. Tai aktualu studentams vykstantiems studijuoti į užsienio universitetus. Savarankiškai mokydami svetimų kalbų ir naudodamiesi kalbos technologijų teikiamomis galimybėmis bei specialiai pateikta mokomąja medžiaga studentai jaučiasi aktyviai dalyvaujantys kalbos įgūdžių lavinimo procese, geriau įsidėmi rūpimus faktus, nes patys stengiasi suprasti tam tikro kalbos fakto vartojimo dėsningumus. Užsienio studentų mokymą lietuvių kalbos kaip svetimos kalbos VGTU universitete yra tyrusi R. Žukienė [26], kuri tyrimo metu atliko studentų naujos kalbos mokymosi poreikių analizę, nustatė, jog efektyviausias kalbos mokymo(si) būdas yra įprastinis mokymas(is) klasėje su mokytoju. Mažiausiai naudos teikiantis būdas – el. medžiaga (kompaktiniai diskai, garso ir vaizdo medžiaga). Reikšmingų darbų stengiantis mokyti užsienio studentus lietuvių kalbos yra pateikusi J. Džežulskienė: „Kalbu lietuviškai“, „Klausau ir skaitau lietuviškai“. Taip pat kalbininkė sutinka, jog mokant kalbos svarbu naudoti įvairias

technologijas. Apie kalbos technologijų taikymą ir ypatumus lietuvių kaip svetimų kalbos mokyme J. Džežulskienė yra parašiusi straipsnį „Kalbos technologijų taikymo lietuvių kaip svetimai kalbai mokytis ypatumai“. Šiame straipsnyje taip pat nurodoma, kad informacinių technologijų naudojimas, kurio pagrindas yra ir nuotolinis / virtualus mokymasis, psichologiniu požiūriu didina besimokančiojo motyvaciją ir pasiekimus [4].

Esant kalbos žinių poreikiui problemą galima spręsti naudojantis metodika bei specialiai tokiai problemai spręsti sukurtu kursu, kuriame bus keliama informacija, padėsianti nesunkiai pramokti lietuvių kalbos pagrindų (bendram suvokimui ir susikalbėjimui), pritaikyta prie informacinių technologijų priemonių kalbų žinioms patikrinti bei kalbos atpažinimo priemonių kalbos tarimo įgūdžiams formuoti.

**Darbo objektas** – užsienio studentų lietuvių kalbos mokymosi procesas.

**Darbo tikslas** – sukurti virtualią lietuvių kalbos mokymosi sistemą panaudojant kalbos atpažinimo technologijas kalbiniams įgūdžiams lavinti.

Tikslui pasiekti reikalinga įgyvendinti šiuos **uždavinius**:

1. Atlikti literatūros šaltinių analizę siekiant teoriškai pagrįsti užsienio kalbų mokymosi ypatybes, metodus ir šnekamosios kalbos atpažinimo, garsynų, gramatikų sudarymo bei transkripcijų temas.
2. Siekiant išsiaiškinti lūkesčius kuriamam kalbos mokymosi kursui, atlikti lietuvių kalbos specialistų, užsienio bei lietuvių studentų apklausas.
3. Apžvelgti esamų kalbos mokymo priemonių mokymo metodines ypatybes ir parengti lietuvių kalbos mokymosi metodiką.
4. Sukurti pavardžių ir vardų garsynus ir atlikti su jais tyrimus.
5. Sukurti lietuvių kalbos mokymo kursą panaudojant kalbos atpažinimo technologijas.
6. Atlikti kurso efektyvumo tyrimą.

Prie darbo pridedamas kurso diegimo aktas. Kursas realiai ketinamas panaudoti KTU univeristete, Socialinių, humanitarinių mokslų ir menų fakultete užsienio studentams mokytis lietuvių kalbos.

**Darbo struktūra.** Darbas susideda iš įvado, analitinės dalies, tyrimų dalies, projektinės dalies, literatūros sąrašo ir 21 priedo. Analitinėje dalyje aprašomas balso atpažinimas, jo pritaikymo metodai, kalbos mokymosi priemonių analizė (apimtis – 19 psl.). Tyrimų dalyje aprašyta kalbos tyrimo eiga, atlikti bandymai ir testai, analizuojami jų metu gauti rezultatai, aprašomas sukurtos atpažinimo programos kūrimo procesas (apimtis – 15 psl.). Projektinėje dalyje aprašytos apklausos bei virtualaus kurso realizavimo principai. Išvadose pateikiamas atliktų darbų apibendrinimas.

## 2. ANALITINĖ DALIS

Siekiant sukurti lietuvių kalbos mokymo kursą užsienio studentams, kuriame būtų remiamasi kalbos atpažinimo technologijomis, šiam tikslui pasiekti reikalinga išanalizuoti užsienio kalbų mokymosi ypatybes, apžvelgti egzistuojančių virtualių kalbos mokymo priemonių galimybes, apibrėžti lietuvių kalbos mokymosi virtualiu būdu galimybes užsienio studentams bei išsiaiškinti kalbos atpažinimo sistemų veikimo principus.

Detalizuojant informacinių technologijų pritaikymą lietuvių kalbai mokyti, skyriuje pateikiama detali skirtingais mokymo metodais besiremianti kalbos priemonių analizė, aptariamas šnekos atpažinimas, jo ištakos, taip pat nusakoma paslėptųjų Markovo modelių (PMM) nauda bei pritaikymas kalbai atpažinti. Analizuojami su tyrimo tematika susiję literatūriniai šaltiniai, apie kalbos atpažinimą bei jam naudojamų garsynų, transkripcijų ir gramatikų sudarymą. Svarbu apžvelgti ankstesnius tyrimus ir rezultatus, gautus atliekant eksperimentus šioje srityje. Literatūrinė medžiaga, į kurią įeina įvairių autorių parašyti darbai panašiomis temomis bei autorių nuomonės yra naudojama baigiamojo magistrinio darbo teorinei ir praktinei dalims pagrįsti ir tyrimo esmei geriau suvokti.

### 2.1. Užsienio kalbų mokymosi ypatybės

Mokant užsienio kalbos dažnai iškyla klausimas, kas šiame procese yra laikomi svarbiausiais aspektais: leksika, gramatika, fonetika ir į ką labiausiai reikia atkreipti dėmesį [14].

Kaip teigia L. Mauzienė, šiuo klausimu nuolat diskutuoja lingvistai ir metodininkai, bet jų nuomonės neretai skiriasi. Vieni teigia, jog svarbiausia yra kalbos garsinė sandara, kiti pirmenybę teikia gramatiniam-struktūriniam kalbos aspektui, tretį (jų dauguma) prioritetiniu dalyku mokant kalbų laiko leksiką. Tai siejama su tuo, kad, mokant 50 žodžių ir tik 5 gramatinių konstrukcijų, galima sudaryti daugiau sakinių, nei žinant penkis žodžius ir 50 gramatinių struktūrų. Vaikas gali išsakyti savo mintis greitai vartodamas atskirus žodžius ir nepaisydamas gramatikos taisyklių. Nepaisant to, artimieji nerišliais sakiniais kalbantį vaiką vis tiek supras. Bendraujant su užsieniečiais gramatinės klaidos nėra kliūtis suprasti pašnekovą. Dažniausiai pritrūkstama žodžių [14].

Kalbą galima traktuoti kaip prasmės, suvokimo perteikimo priemonę, kuri didžiąja dalimi priklauso nuo konteksto [12]. Ugdomam žmogui būtini ir kultūrinės, ir kalbinės komunikacijos gebėjimai, kurie vienas nuo kito neatskiriami mokantis užsienio kalbų, t.y. įgyjant bei ugdant komunikacinę kompetenciją, kuri susideda iš kelių subkompetencijų: lingvistinės, socialinės lingvistinės, diskursinės, strateginės, socialinės kultūrinės, socialinės.

„Ugdantis lingvistinę subkompetenciją būtina pažinti ir įvaldyti kalbos struktūros dėsningumus, sukaupti žodyną, mokėti abėcėlės raides, garsus bei suprasti jų santykius, gebėti pakankamai taisyklingai tarti, kirčiuoti, intonuoti bei skirstyti sakinius prasminėmis grupėmis“ [11].

Taip pat mokomasi naudotis žodynais ir gramatikos žinynais. Leksikos išmokstama skaitant įvairius tekstus. Įgyjamas bendriausias supratimas apie kalbos funkcionavimą.

**Ugdantis socialinę lingvistinę subkompetenciją** mokomasi interpretuoti kalbos formas pagal tam tikras situacijas, taip pat jas vartoti. Kiekviena situacija nulemia kalbos pasirinkimo formą.

**Ugdantis diskursinę subkompetenciją** mokomasi formuluoti rišlias žodžių frazes, sakinius ir tekstus, juos suprasti [11], o **ugdantis strateginę subkompetenciją** mokomasi naudotis įvairiais metodais tam, kad būtų lengviau suprasti skaitomą ar klausomą tekstą.

„**Ugdantis socialinę kultūrinę subkompetenciją** siekiama suvokti socialinį kultūrinį kontekstą, kuriame kalba vartojama“ [11]. Šios subkompetencijos ugdymas vyksta visais kalbos mokymo(si) būdais, naudojant įvairius šaltinius (pavyzdžiui, skaitant ir klausant informacinių tekstų, garsinėmis bei vaizdinėmis priemonėmis ir pan.). Taip pat susipažįstama su kultūrinėmis sritimis: kasdienio gyvenimo savybėmis, tradicijomis, papročiais, kitais kultūros elementais (pavyzdžiui, daile, architektūra, muzika), istorija, geografijos ir kitų sričių faktais.

**Puoselėjant socialinę subkompetenciją** išugdomas pasitikėjimas savimi, tinkamas elgesys socialinėse situacijose bei pasitikėjimas savimi ir noras bendrauti su kitais [11].

Taigi siekiant užtikrinti gerus išmoktos užsienio kalbos įgūdžius labai svarbu stengtis įvairiais būdais praplėsti vartotojo žodyną įvairiuose kontekstuose. Turi būti lavinami rašymo svetima kalba įgūdžiai, nepamirštant gramatikos tobulinimo. Labai tikslinga bendrauti su kitataučiais, kurie užtikrintų galimybę spontaniškiems pokalbiams. Praktinės užduotys turėtų būti pateikiamos su nuolat pasikartojančia įsisavinta medžiaga. Tačiau svarbiausia mokymosi procese motyvacija, dėl kurios mokymosi procesas tampa tarsi pramoga, o žinias tampa nesunku įsisavinti.

## **2.2. Virtualios lietuvių kalbos mokymosi priemonės**

Vykstant informacinės visuomenės plėtrai, sukūrus skaitmeninių priemonių, skirtų kalbos vienetams kaupti, tirti ir lyginti, žodynams sudaryti, automatinėms vertimo priemonėms rasti, ryškėja poreikis visas šias priemones vienu ar kitu būdu integruoti į užsienio kalbų mokymo procesą.

Pasak J. Mikėlionienės, lietuvių kalbai mokyti(s) yra sukurta įvairių mokymo(si) programų, įrankių ir priemonių. Kai kurios iš priemonių yra skirtos gimtakalbiams, kitos – pradedantiems mokytis lietuvių kaip svetimos kalbos arba jau pažengusiems ir norintiems tobulinti rašybos, skyrybos, kirčiavimo bei bendravimo įgūdžius ir pan. Tokio tipo programos gali būti pateikiamos internete (pvz., „Po truputį“, „Eksternas“, „Kieti riešutėliai“ ar net moksleivių sukurta interneto svetainė „Klaidutis“) arba pateikiamos kompaktiniuose diskuose („Žodis po žodžio“), vadovėliais (J. Džežulskienės „Lietuvių kalba kitakalbiams“). Toks netradicinis ir kompiuterinėmis technologijomis paremtas kalbos mokymas(is) skatina besimokančiuosius domėtis kalba [15].

Mokymuisi naudojamos IT priemonės turi nemažai galimybių, galinčių mokymosi turinį pateikti garso, vaizdo ar rašto pavidalu. Geru pavyzdžiu galima laikyti kalbos mokymo(si) priemones

„SurfaceLanguages“ ir „Loecsen“. Ieškant internete straipsnių apie šias priemones informacijos randama nedaug.

„SurfaceLanguages“ yra tinklalapis, sukurtas 2009 metais asmens, pasivadinusio Moon Face. Šis tinklalapis buvo įkurtas siekiant kilnaus tikslo – surinkti frazes ir kitus kalbos vienetus iš visų egzistuojančių šalių ir patalpinti juos šiame <http://www.surfacelanguages.com/> tinklalapyje. Interneto svetainėje galima išmokti pagrindinių frazių net 42 kalbomis, t.y. ispanų, prancūzų, rusų, danų ir kitomis, tarp kurių galima aptikti ir lietuvių kalbą. Kadangi apie tinklalapį nėra pateikta daug informacijos, o kūrėjas prisistato slapyvardžiu, todėl galima manyti, jog svetainės kūrėjas yra savamokslis kalbų tyrinėtojas, kurio indėlis į kalbų mokymuisi skirtų naudingų priemonių kūrimą didžiulis.

Kita priemonė, kuri gali būti naudojama mokinantis kalbos yra „Loecsen“. Informacijos apie tai, kur ir kieno sukurta ši priemonė nėra, tačiau naudojantis šia nemokama internete prieinama priemone galima išmokti pagrindinių kalbos frazių 32 kalbomis, tokiomis kaip: vokiečių, anglų, ispanų, lietuvių ir kitomis. Šios internetinės kalbos mokymo priemonės tikslas – pateikti svarbiausią kalbinę informaciją, kuri būtų naudinga bendraujant su vietiniais žmonėmis užsienio šalyse. Svetainėje pateikiamas kalbos mokymosi būdas bei suformuota mokymo strategija veda prie užsibrėžto tikslo – kalbos žinių.

Taigi minėtosios priemonės yra tarsi pradžiamoksliai, kurių teikiama medžiaga apie kalbą yra pakankama asmeniui, siekiančiam įgyti pagrindinių kalbos žinių, kurios tikrai pravers išvykus į užsienio valstybę. Kokias kalbos žinias naudojantis šiomis priemonėmis galima įgyti bus apibrėžta panagrinėjus kiekvieną įrankį plačiau.

### **2.2.1. Priemonių SurfaceLanguages ir Loecsen apžvalga**

Priemonės „SurfaceLanguages“ ir „Loecsen“ yra priemonės, paremtos skirtingais kalbų mokymo būdais. Kiekviena priemonė bus apžvelgta atskirai ir nurodoma, kokie būdai yra naudojami siekiant pateikti kalbines žinias šioms įtvirtinti.

Pirmiausia patekus į „SurfaceLanguages“ priemonės tinklalapį vartotojas supažindinamas su pirmine informacija: kokių su kalba susijusių dalykų galima išmokti tinklalapyje, kokia kalba tai bus daroma ir pan. Toliau galima pereiti prie kalbinių dalykų, iš kurių pirmasis – pageidautinos išmokti kalbos pasirinkimas, šiuo atveju, tai yra lietuvių kalbos, nes reikia įvertinti pateikiamos informacijos reikšmę, ar ji vertinga, ar gali praversti užsieniečiams, jeigu šie pageidautų pramokti kalbos naudodamiesi šia kalbos priemone. Pasirinkus pageidaujamą kalbą patenkama į kitą svetainės langą, kuriame tikslingai pateikiama informacija, suteikianti pirmąjį įspūdį apie pasirinktą kalbą, kalbai mokytis naudojamus spalvų, vaisių, mėnesių ir daržovių išmokymo žaidimus, taip pat keletas kūrėjo įžvalgų apie pačią šalį, kalbos ypatybes. Pirmiausia svetainėje siūloma, o ir tikslinga būtų susipažinti su pagrindinėmis mandagumo frazėmis. Jos pateikiamos lentelė anglų (daugumai žmonių žinoma

kalba) ir pageidautina mokymuisi – lietuvių kalba. Angliška frazė *hello/hi* supažindina su pirmuoju žodžiu, toliau seka lietuviškas atitikmuo *labas* ir šis galiausiai dar ištiriamas diktoriaus, nes negirdėdamas žodžių tarimo niekada negali būti tikras, ar gerai žodis bus perskaitytas ir ar gerai pats jį vėliau tarsi. Sekantis žodžių žaidimas yra tarsi savotiška rašybos praktika, susijusi su žodžių formavimu. Viename stulpelyje pateikiami žodžiai anglų, kitame – lietuvių kalba, o žemiau pateikiami dar ir lietuviško stulpelio žodžio skiemenys, iš kurių reikia sudaryti žodžius. Gerai sudarius žodį, jis yra ištiriamas ir žaidimas žaidžiamas tol, kol nelieka laisvų skiemenų, o lietuviškų žodžių stulpelyje esantys žodžiai šviečia stambiu žaliu šriftu. Susipažinus su paprasčiausiais žodžiais kartojami mokymai susiję su sudėtingesniais dalykais, tokiais kaip savęs pristatymas, trumpi klausimai, mokantys užmegzti pokalbį apie sportą, frazės, naudojamos pokalbiui parduotuvėje, viešbutyje. Taip pat nepamirštama leksika, susijusi su pinigais, sveikata, kūno dalimis ir pan. Žaidimai, lavinantys kalbos įgūdžius mokantis spalvų, vaisių, mėnesių ar daržovių pavadinimų yra paremti kitokiu mokymo metodu nei pagrindinių frazių mokymas. Čia, pasirinkus vieną iš žaidimų, matomi vaisių ar daržovių vaizdai ir kiekvienas iš jų tariamas. Išklausus diktoriaus reikia pasirinkti teisingą paveikslėlį, renkamasi tol, kol paspaudžiamas teisingas variantas. Šiame žaidime žymimas kiekis, kiek iš pateiktų variantų buvo pasirinkti teisingai. Tai yra tarsi paskatinimas atsiminti, nes žinios vertinamos balais.

„Loecsen“ kalbos mokymo priemonė remiasi kitokia kalbos mokymo metodika bei informacijos išdėstymu tinklalapyje. Patekus į „Loecsen“ svetainę iš karto matomas žemėlapis su tam tikrų valstybių vėliavomis. Tai interaktyvus „Google street view“ žemėlapis su vėliavomis tų šalių, kurių kalbos galima pramokti naudojantis šia priemone.

Po žemėlapiu pateikiamas kalbų sąrašas, iš kurio išsirenkama lietuvių kalba. Pastebimas dalykas, kurio nėra „SurfaceLanguages“ svetainėje, yra tas, kad lietuvių kalbos mokomasi frazes skaitant nebūtinai anglų kalba. Vartotojų patogumui galima mokytis lietuvių (ar kitos) kalbos skaitant frazes savo gimtąja kalba, pvz.: jei vartotojas yra Čekijos pilietis, gerai mokantis anglų kalbą, jis gali pasirinkti mokytis lietuvių kalbą ir matyti frazes čekų ir anglų kalba. Pirmosios priemonės atveju frazes buvo galima matyti tik anglų kalba ir mokomąja lietuvių kalba. Nusistačius kalbą ir pradėdant mokymus čia, kaip ir pirmuoju atveju, mokomasi mandagumo frazių, tačiau jos ne tik diktuojamos diktoriaus paspaudus ant kiekvieno lietuviško atitikmens, bet matomas ir paveikslėlis, iliustruojantis frazę, o po juo ir gimtosios ar kitos vartotojui suprantamos kalbos vertimas. Dar vienas naudingas dalykas pastebėtas šioje priemonėje – prie tam tikrų išsireiškimų pridėtas dar vienas kalbos raiškos variantas, pvz. šalia frazės *kiek tai kainuoja* (kurią yra naudinga žinoti, bet tikras Lietuvos pilietis vargu ar tokiu būdu klausia pardavėjo prekės kainos) pridedama ir labiau realioje situacijoje vartojama frazė *kokia kaina?* Susipažinus su mandagumo frazėmis svetainėje galima rasti ir pokalbiui užmegzti reikalingų frazių su atsakymais: *ką tu čia veiki?*

*Atostogauju*. Pastebima, kad šioje svetainėje mokymas labiau orientuojamas ne į leksiką, bet į konkrečias frazes, todėl, kad pereinant į *jausmų* stulpelį nematome pateikiamo žodyno, nes matomos suformuotos frazės tokios kaip: *aš esu laimingas, man karšta*. Gerai yra tai, kad jeigu frazė tinka tik vyriškosios giminės atstovams nurodoma ir moteriškoji frazės versija *aš esu laiminga*. Suprantama, kad pramokus frazių tokiomis temomis kaip: jausmai, spalvos, pokalbiai bare ir pan. galima atlikti žinių patikrinimą. Žinių patikrinimo metu diktorius ištaria frazę lietuviškai ir iš stulpelyje pateiktų frazių gimtąja kalba reikia pasirinkti teisingą variantą. Rezultatas – mokymosi progresas skaičiuojamas balais.

Nors šias internetines priemones sieja bendras tikslas – pateikti medžiagą, kurią išmokęs užsienietis galėtų susikalbėti su Lietuvos piliečiais ir suvoktų bei galėtų kalbėtis keliomis skirtingomis temomis, tačiau žinių gilinimas yra neišvengiamas. Mokymosi progresas būtų greitas, jeigu besimokantysis dalyvautų pokalbiuose (internetiniuose taip pat) su tikrais lietuviais, kas leistų patobulinti šnekamosios kalbos žinias. Naudinga būtų tobulinti ir tarimą. Tam reikalinga programa, galinti atpažinti vartotojo ištartą žodį ir jį palyginti su esančiu duomenų bazėje bei duoti atsaką, ar šis buvo ištartas teisingai.

### **2.2.2. Kalbos mokymosi priemonė Duolingo su integruota kalbos atpažinimo funkcija**

2.2.1. – amė poskyryje buvo aptartos „SurfaceLanguages“ ir „Loecsen“ kalbos priemonių galimybės. Tačiau kalbos mokymasis minėtomis priemonėmis nėra pilnavertis mokymosi procesas, nes nėra visapusiškai išpildytas lingvistinės subkompetencijos įsisavinimas. Taip yra todėl, kad priemonėse nėra galimybės mokytis tarti kalbos garsus, formuoti sakinius žodžiu ir pan. Taip yra todėl, kad priemonėse nėra įdiegta kalbos atpažinimo funkcija, leidžianti pasitikrinti įgytas kalbos žinias tariant jas žodžiu. Tokią funkciją atlieka priemonė „Duolingo“, tačiau ši neturi galimybės mokytis lietuvių kalbos, tačiau jos mokymo metodiką taip pat svarbu apžvelgti.

„Duolingo“ yra kalbos mokymo(si) terpė, įdiegta internetiniame tinklalapyje. Ši priemonė yra sukurta Pitsburge (JAV) 2011 metais IT specialistų Luis von Ahn ir Severin Hacker. Jie nurodė, kad pagrindinė „Duolingo“ priemonės esmė – užsienio kalbos mokymasis bei pagalba kitiems verčiant tekstus. Naudojant šią nemokamą internete prieinamą priemonę galima išmokti 20 skirtingų užsienio kalbų, pvz: ispanų, portugalų, vokiečių, lenkų, ukrainiečių ir kitų. Šios internetinės kalbos mokymosi priemonės tikslas – suteikti galimybę įgyti aukštos kokybės pasirinktos kalbos žinių. Naudojantis „Duolingo“ priemone galima išmokti pagrindinių kalbos frazių, parinkti tinkamą žodį sakiniui, išmokti tarti žodžius, juos parašyti ir pan. Pasirinktos kalbos 34 valandų kursas yra prilyginamas kalbos mokymosi pusmečiui universitete.

Mokantis kalbos naudojant „Duolingo“ pravartu mokėti anglų kalbą, nes ši kalba yra pagrindas, ją pasitelkus mokoma svetimų kalbų, nes pačios anglų kalbos šiame tinklalapyje išmokti nėra galimybės, kaip minėta anksčiau, lietuvių kalbos taip pat. Mokymai prasideda nuo lengviausio



*Basics1* – pradinio lygio. Jame yra kelių tipų užduotys. Viena iš jų – surasti atitikmenis tarp žodžio ir vaizdo. Ekrane matomi 3 vaizdai ir po jais esantys vaizdus nusakantys pavadinimai. Virš vaizdų – žodis, kuriam reikia priskirti atitikmenį. Pele reikia paspausti ant žodį atitinkančio vaizdo. Taip tikrinamos elementarios žinios. Derinys – vaizdas ir žodis leidžia geriau įsiminti informaciją. Paspaudus ant kiekvieno iš paveikslėlių išgirstama, kaip vienas ar kitas žodis tariamas. Gerai atlikus užduotį pereinama prie kitos, blogai – sistema fiksuoja klaidą, kurią galima, bet nebūtina ištaisyti pasirinkus tinkamą variantą. Atlikus keletą tokio tipo užduočių sistema iš karto „sugalvoja” patikrinti jau išmoktas žinias. Besimokantysis turi užrašyti išgirsto žodžio vertimą. Užduotys kombinuojamos. Nuo pat mokymosi pradžios taikomi skirtingi mokymosi metodai. Dar vienas priemonėje taikomas mokymo metodas – trūkstamo žodžio parinkimas į reikiamą vietą arba teisingas sakinio vertimo variantas, kurį reikia pasirinkti. Sistemos kūrėjai priemonėje integravo kalbos atpažinimo funkciją ir pateikė naudotojui užduočių: versti žodžius arba sakinius arba juos pakartoti pagal tai, kaip taria sistema, žodžiu. Blogai ištarus arba išvertus sakinį reikia užduotį kartoti. Jeigu sakinys išverčiamas tik iš dalies gerai arba nepavyksta to niekaip atlikti, grįžtama prie ankstesnių užduočių, taip uždraudžiant tęsti mokymosi procesą, kol visos pradžioje pateiktos žinios nebus įsisavintos. Taip pat yra ir klausymo užduočių: balsas taria žodį arba sakinį, o besimokantysis privalo užrašyti nurodytame lange tai, ką girdi. Atsižvelgiama į besimokančiojo lygį, todėl pradiniam etape sakinys tariamas lėtai.

Atlikus visas užduotis ir perėjus prie sunkesnio mokymosi lygmens, užduočių metodų taikymas išlieka panašus, bet užduotys tampa sudėtingesnės: tariami sunkesni žodžiai, reikia išversti sunkesnius sakinius, žodžius ir pan. Ši kalbos mokymo sistema yra paruošta mokyti kalbos visapusiškai: tarimo, leksikos, gramatikos ir pan. Sistemoje prie kiekvienos naujos temos yra trumpi aprašymai apie kalbą: vyrišką, moterišką gimines, jų vartojimą su pavyzdžiais, kurie vyraus einamosiose temose.

„Duolingo“ priemonėje yra galimybė mokantis kalbos lavinti ir kalbines žinias verčiant sakinius bei kitus leksikos vienetus žodžiu, kai yra naudojama kalbos atpažinimo funkcija (priemonės „SurfaceLanguages“ ir „Loecsen“ to negali atlikti), bet nėra galimybės išmokti lietuvių kalbos, kai internete galima rasti priemonių, kurios leidžia mokytis lietuvių kalbos, tačiau šios neturi kalbos atpažinimo funkcijos. Taigi žinant, kokių kalbos ypatybių privalu įgyti mokantis kalbos (sužinota 2.1. – amė poskyryje) bei kaip šios gali būti pateikiamos (įvairių užduočių pavidalu) galima susidaryti bendrą vaizdą, kaip atrodys lietuvių kalbos kursas, skirtas užsieniečiams su jame pateikta priemone, kuri leidžia lavinti kalbos įgūdžius žodžiu naudojant programėlę, paremtą kalbos atpažinimu. Dėl minėtųjų privalumų „Duolingo“ priemonę taip pat galima laikyti tinkama kalboms mokytis, nes jose galima įgyti visų kalbinių įgūdžių.

### **2.2.3. Oneness kalbų mokymo kursas**

Internetė galima rasti įvairių, tokių kaip minėtieji „SurfaceLanguages“, „Loecsen“ arba „Duolingo“, kurie yra įvairių korporacijų, kalbos mylėtojų ir pan. sukurti produktai. Tačiau gausu ir internetinių kursų, kurie remiasi kitokiu kalbos mokymo principu nei įrankiai ir yra sukurti lingvistų.

Vieni iš tokių, Vilniaus universitete sukurti estų, suomių, lenkų, portugalų ir lietuvių kalbų internetiniai kursai, pavadinimu „Oneness“. Tai šiuolaikiški, nemokami, patogūs, ir prieinami kiekvienam vartotojui kompiuteriniai kursai. Kursų tikslinės grupės yra mainų programų studentai bei migrantai. Kursai skiriami žmonėms, kurie nori pasiekti pasirinktos kalbos A1 mokėjimo lygį. Pagrindinis kursų didaktinis metodas – komunikacinis kalbos mokymas. Pagrindinis mokymo ir mokymosi tikslas – ugdyti komunikacinę kompetenciją, t. y. mokytis vartoti pasirinktą kalbą įprastose gyvenimo situacijose, kai reikia klausyti, skaityti, kalbėti ir rašyti.

Sukurti kursai nėra akademiški, juose daug įdomių užduočių, pvz. testas – kompiuterinis kalbinis žaidimas, imituojantis realią kalbų vartoseną. Kursą galima rasti interneto svetainėje <http://www.oneness.vu.lt>. Šio lietuvių kalbos kurso trūkumas – nelavina kalbinių įgūdžių naudojant kalbos atpažinimo technologijas.

Nepaisant minėtojo trūkumo, „Oneness“ kalbos mokymosi kursai sukurti lingvistų, todėl jame pateikta mokymosi medžiaga pertekia reikiamą kalbinę informaciją vartotojams, aiškiai nusako kalbines panaudojimo sritis ir subtilybes [16].

### **2.2.4. Lietuvių kalbos mokymosi virtualiu būdu galimybės ir rekomendacijos užsienio studentams**

Kiekvienas besiruošiantis kalbą mokytis žmogus žino, kokių tikslų ketinama įgyti užsienio kalbos žinių, todėl tai, kokios kalbos žinios jam bus reikalingos, jis taip pat gali suvokti pats. Jeigu kalbėtume apie tokį atvejį, kai užsienio studentas ketina įgyti svetimos kalbos žinių dėl to, kad nuvažiuotų studijuoti į svetimą šalį galėtų susikalbėti su vietiniais žmonėmis, tai išvada būtų ta, jog pirmiausia jam bus reikalingi kalbos pradmenys – elementariausias kalbos suvokimas. Tai reiškia, kad remiantis 2.1 – amė poskyryje pateikiama informacija asmeniui yra reikalinga: ugdyti lingvistinę kompetenciją, kad būtų išmokta lietuvių kalbos abėcėlė, kaupiamas žodynas, stengiamasi taisyklingai tarti žodžius bei bandoma suprasti kalbos struktūros dėsnumus. Ugdytis socialinę lingvistinę kompetenciją taip pat svarbu, nes mokomasi vartoti kalbos formas pagal situaciją, o jų pasitaiko daug ir įvairių. Šių dalykų laisvai internete prieinamos bei nemokamos aptartosios kalbos priemonės gali suteikti kiekvienam norinčiam išmokti lietuvių kalbos. Tiesa, abėcėlė nė vienoje iš priemonių nėra akcentuojama, todėl šią sritį rekomenduotina patobulinti. Diskursinės subkompetencijos ugdymas naudojantis virtualiomis mokymosi priemonėmis yra įmanomas, nes formuluoti tekstus ir pasakymus nėra sunku, ypač iš sukauptų pavyzdžių arba naudojantis žodynu. Strateginė subkompetencija internetinėse priemonėse pasireiškia tuo, jog kalbos mokoma naudojant skirtingus būdus – strategijas, todėl tai leidžia suprasti mokomąją medžiagą. Socialinė kultūrinė

subkompetencija, kuria siekiama suvokti kalbos vartojimą socialiniame kulturniame kontekste, nėra pateikiama (užduočių ar kitokios medžiagos pavidalu) priemonėse. Tai galima laikyti priemonių trūkumu, nes suvokti, šalies kultūros sritis (meną, techniką, ūkį) yra ne tik naudinga, bet ir įdomu. Galiausiai socialinės subkompetencijos puoselėjimas, juo yra išugdomas žmogaus mokėjimas tinkamai elgtis įvairiose socialinėse situacijose, pasitikėjimas savimi bei noras bendrauti su kitais žmonėmis. Visa tai yra reikalinga, tačiau priemonės to neskatina jokiais užduotimis. Tai priklauso nuo žmogaus būdo [14].

Remiantis analize, galima teigti, kad priemonės gali suteikti besimokančiajam pirminių ir svarbiausių žinių, kurių užsienio studentui gali prireikti kasdieninėse situacijose esant svetimose šalyje (šiuo atveju Lietuvoje), todėl išbandyti šias priemones naudinga ne tik studentams, bet visiems, norintiems įgyti kalbos pagrindų.

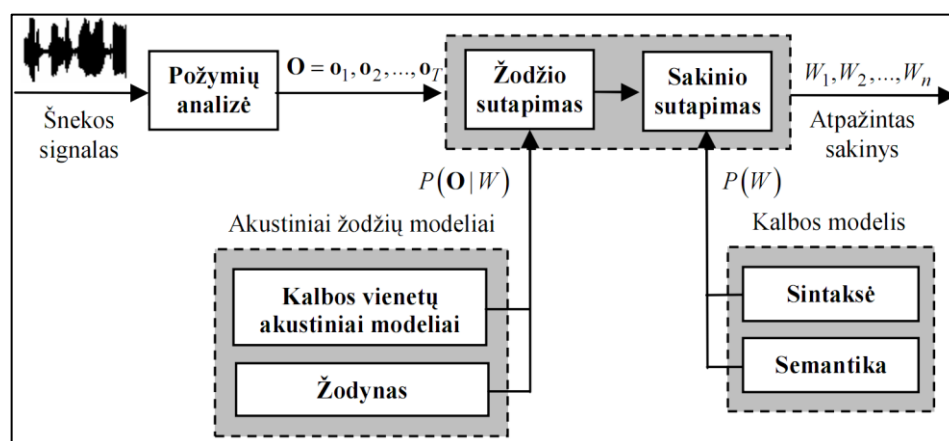
Minėtieji įrankiai yra laisvai prieinami internete ir jais, turint poreikį, galima naudotis nemokamai. Tokio tipo kalbos priemonės naudinga reklamuoti dažnai žmonių lankomose interneto svetainėse, taip pat ir universitetų tinklalapiuose prie tam tikrų specialybių, orientuojant į užsienio studentus, greta nurodant, jog reikia turėti bent menkiausias kalbos žinias, siekiant įstoti į tam tikrą specialybę. Prierašus su įvardinta sąlyga galima išskirti dėmesį patraukiančiu skiriamuoju ženklu, kuriame būtų užkoduota nuoroda su konkrečiu interneto adresu arba nurodymu, vedančiu prie naudingų kalbos priemonių, padėsiančių išspręsti kalbos nemokėjimo problemą.

### **2.3. Automatinis šnekamosios kalbos atpažinimas**

Iš esmės balso atpažinimo technologija yra ne naujas dalykas. Reikšmingą indėlį kalbos atpažinimui davė 1877 metais Tomo Edisono išrastas fonografas. Tai pirmasis prietaisas, skirtas garsui įrašyti ir atkurti. Pasak internetinių šaltinių, Tomas Edisonas buvo sudaręs netgi 10 – ties dalykų sąrašą, kur, pasak jo, būtų galima pritaikyti fonografa: diktavimui, kalbančioms knygelėms, muzikai įrašyti, šeimos albumas įgarsinti, kalbantiems žaislams ir laikrodžiams, muzikinėms dėžutėms, žmonių sakomoms kalbos įrašyti, aklųjų gyvenimui palengvinti, mokymosi tikslais, kaip pagalbinį prietaisą telefonui. Fonografo sukūrimas, jo pritaikymo idėja, galimybės atlikti veiksmus su garsų atkūrimu ir sąrašas, apibūdinantis kalbos atpažinimą ir pritaikymą nurodančias terpes mokslininkams galėjo tapti stimulu bandyti sukurti pirmiausia paprastus išradimus, galinčius bendrauti su žmogumi, o vėliau progresuoti iki išmaniųjų technologijų lygio. Manoma, kad kalbos atpažinimo istorija prasidėjo 20 amžiuje – 1920 metais su mažo žaisliuko Radio Rex sukūrimu. Žaisliukas reagavo į ištartą žodį „Rex“. Iš ceciuoloido pagamintas šuo ištarus jo vardą „Rex“, „iššokdavo“ iš savo būdos. Žaislo-prietaiso veikimas yra paremtas jo pagrinde įtaisyto šunto jautrumu 500 Hz akustinei energijai (ja pasižymėjo žodžio „Rex“ balsė). Minėtąja energija paveikus šuntus, srovė, maitinanti pagrinde įtaisytą magnetą, nutrūkdavo ir šuo, veikiamas suspaustos spyruoklės, „išbėgdavo“ iš būdos. Žinoma, žaislinis šuo reagavo ir į kitus žodžius ir netgi į

atsitiktinius garsus, kurių spektre buvo pakankamo lygio 500 Hz dažnio dedamoji. Nors Radio Rex atrodo kaip labai paprasta sistema, tačiau joje įgyvendintas dabartinėse sistemose naudojamas atpažinimo principas: etaloninio garso savybės išsaugojamos ir lyginamos su nagrinėjamoju garsu (etaloninio garso savybes atitiko tam tikram dažniui jautrus šuntas) [23]. Pirmieji automatinio kalbos atpažinimo bandymai buvo atlikti 1952 metais Belo laboratorijoje. Tai buvo izoliuotų žodžių vienam kalbėtojai sistema. Taip pat reikšmingas indėlis šiai technologijai padėtas IBM korporacijos ir Carnegie Mellon universiteto mokslininkų 8-ojo dešimtmečio pradžioje. Balso apdorojimo technologijos sritis apima įvairius mokslus: taikomąją matematiką, lingvistiką bei informatiką ir kt. Nuo to laiko balso apdorojimo sritis ėmė plėstis. Tyrėjai iš įvairių kompanijų ir universitetų stengėsi, sukurti sėkmingą balso dialogo sistemą, kuri galėtų interaktyviai bendrauti su žmogumi, siekiant pagreitinti informacijos apdorojimą kompiuteriu ir kitais informacijos apdorojimo įrenginiais.

Šnekos atpažinimas dar žinomas kaip automatinis kalbos atpažinimas – tai procesas, atpažįstantis ir interpretuojantis žmogaus kalbą, ją paversdamas kompiuteriui suprantama signalų kalba [5]. Jurafsky pateikia dar tikslensnį kalbos atpažinimo apibrėžimą: automatinis kalbos atpažinimas – tai kompiuterio sistemos darbo procesas, kuris atlieka žodžių atitikmenų akustiniams signalams parinkimą [7]. Pagrindinis automatinio kalbos atpažinimo sistemų kūrimo ir tobulinimo tikslas yra technologijas, kurios kaip ir žmogus galėtų girdėti, kalbėti bei suprasti. Taip pat veikti pagal balsu gautą informaciją. Automatinis kalbos atpažinimas yra pirmoji grandis balsinių technologijų produktuose teigia Laurinčiukaitė [10].



**1 pav.** Automatinio kalbos atpažinimo sistemos struktūra

Nors iš išorės gali atrodyti, kad automatinis kalbos atpažinimas yra tiesiog paprastas procesas, tačiau iš tikrųjų tai – sudėtingų veiksmų seka. Šnekos atpažinimo procesas prasideda šnekos signalo įvedimu į automatinio šnekos atpažinimo sistemą ir baigiasi tekstinio atitikmens akustiniu pateikimu. Kaip teigia H. Pranevičius, Š. Raudys ir kt. [17], nesvarbu, kokia būtų automatinio kalbos atpažinimo taikymo sritis, visada galima išskirti dvi automatinio kalbos atpažinimo sistemos funkcines dalis: pirminio apdorojimo bloką (*front – end*) ir antrinio apdorojimo bloką (*back – end*). „**Pirminio**

**apdoravimo bloko paskirtis** – priimti šnekos signalo diskretas, jas suskaidyti nustatyto dydžio intervalais, atlikti šio intervalo analizę vienu iš pasirinktų analizės metodų ir, prireikus, atlikti kokias nors papildomas signalo transformacijas. Pirminio apdoravimo bloko išėjime suformuojamas tam tikras tam tikrą šnekos signalo fragmentą aprašančių požymių vektorius. **Antrinio apdoravimo bloko paskirtis** – priimti suformuotų požymių vektorių seką ir, naudojantis vienu iš statistinio atpažinimo algoritmų, priskirti požymių vektorių vienai iš galimų klasių. Priskyrimas visuomet atliekamas lyginant požymių vektorių su tam tikrais etaloniniais požymių vektorių pavyzdžiais, kurie buvo nustatyti anksčiau“ [17]. Kiti statistinio klasifikavimo uždaviniai nuo automatinio šnekamosios kalbos atpažinimo išsiskiria tuo, jog tenka klasifikuoti nevienodo dydžio požymių rinkinius, nes skirtingi žodžiai yra nevienodos trukmės, juos sudaro skirtingas fonemų skaičius, skirtingi kalbėtojai pasižymi nevienodu kalbėjimo tempu, tempas skiriasi nuo emocinės bei psichinės būsenos ir kitų panašių veiksnių.

Dar vis kalbant apie automatinių kalbos atpažinimą, Airenas Vaičiūnas [25] teigia, jog dažniausiai skiriamos keturios pagrindinės automatinio kalbos atpažinimo dalys:

**Požymių išskyrimas.** Šioje dalyje iš garso signalo išskiriami požymiai  $O$ , kurie bus naudojami tolesniame atpažinime.

**Akustinis modelis.** Tai modelis, kuris turi įvertinti tikimybę  $P(O/W)$ .

**Kalbos modelis.** Jo paskirtis yra įvertinti *a priori* tikimybes  $P(W)$ .

**Paieška hipotezių erdvėje.** Apima akustinio ir kalbos modelių sujungimą bei paiešką žodžių hipotezių erdvėje.

Visgi kalbos atpažinimui modeliuoti yra taikomi įvairūs metodai, kuriams skirtingi modeliai. Vienas iš labiausiai taikomų modelių yra PMM (paslėptieji Markovo modeliai), apie kuriuos šiame darbe bus kalbama vėliau ir pasakoma nemažai svarbių dalykų, nes šių modelių taikymas yra svarbus atliekant baigiamojo magistro darbo praktinę dalį.

Anot G. Tamulevičiaus [23], yra trys pagrindiniai atpažinimo sistemų darbo režimo parametrai, pagal juos galima klasifikuoti kalbos atpažinimo sistemas:

- **Atpažinimo objektas.** Priklausomai nuo to, kokiame lingvistiniame lygmenyje vyksta kalbos atpažinimas, sistemas galima skirti į du tipus: pavienių žodžių atpažinimo ir ištisinės kalbos atpažinimo. Pirmuoju atveju, analizuojami atvejai yra pavieniai žodžiai atskirti tyła. Antrasis atvejis – sistemos, kurios nagrinėja kalbą, neturinčią tylos intarpų. Tokia kalba yra ištisinė analizuojama ištisinės kalbos atpažintuvais. Tai gali būti skaitymas balsu, sakytinė arba spontaniška kalba ir kt.
- **Žodynas.** Žodynu vadiname visumą skirtingų lingvistinių vienetų, kuriais yra apmokyta sistema (sukurti etalonai). Žodyno dydis yra svarbus kalbos atpažinimo sistemos parametras, nes klaida atpažinimo procese dažnai įvyksta, kai bandoma atpažinti žodyne nesantį žodį.

Atsižvelgiant į žodyno dydį, atpažinimo sistemos gali būti skirstomos į mažo (šimtų etalonų eilės), vidutinio (tūkstančio eilės), didelio (dešimčių tūkstančių eilės) ir labai didelio (šimtų tūkstančių ir daugiau) žodyno sistemas.

- **Priklausomybės nuo kalbėtojo.** Pagal jautrumą kalbančiajam atpažinimo sistemos gali būti skirstomos į priklausomas nuo kalbančiojo ir nepriklausomas. Priklausomoje nuo kalbančiojo sistemoje gali būti nagrinėjami tik sistemą apmokiusiojo asmens ištarti žodžiai. Jeigu norima, kad tokia sistema atpažintų kito asmens kalbą, reikia sistemą mokyti iš naujo, kito asmens tariamais žodžiais.

Greta išvardintųjų parametru, kalbos atpažinimo sistemos gali būti klasifikuojamos pagal tai, kaip kalbos sistemoje yra realizuotas atpažinimo metodas, signalo analizės metodas, ar netgi atpažįstama kalba (pvz., anglų, lietuvių) [23].

Pasak Vaičiūno [25], šiuo metu vis daugiau žmonių dirba su kompiuteriais ar kitais informacijos apdorojimo įrenginiais. Didelė dalis laiko skiriama informacijai surinkti kompiuteriu. Galimybė šią užduotį atlikti balsu labai sumažintų darbo laiką, kurį žmogus praleidžia surinkdamas tekstą naudojantis klaviatūra ir pele. Dėl šios priežasties, pasaulyje, spręsdami minėtąją problemą dirbuojasi kalbos atpažinimo specialistai, nes pasaulyje plačiai paplitusios kalbos, tokioms kaip anglų, ispanų ir kt., jau yra sukurtos automatinio kalbos atpažinimo sistemos, todėl reikia įdėti daug pastangų, kad ir lietuvių kalba turėtų tokią sistemą. Tai nereiškia, jog lietuviai nesistengia įgyvendinti panašių tikslų. Lietuvių kalbos atpažinimo tyrimai taip pat yra vykdomi Lietuvos mokslo įstaigose, todėl galima pasidžiaugti bent tuo, jog yra sukurtos bent eksperimentinės kalbos atpažinimo sistemos. Žinoma, yra keletas paskelbtų darbų, susijusių su rišlios kalbos atpažinimo modeliais, tačiau labai didelio žodyno rišlios šnekos atpažinimo problema dar nebuvo nagrinėta. Tokioms kalbos atpažinimo sistemoms reikalingas kalbos modelis, o tokiems tyrimams reikalingi tekstynai, atsirado tik prieš keletą metų.

Automatinis kalbos atpažinimas gali būti pritaikomas įvairiose žmonių veiklos srityse, bet keletas svarbesnių demonstracinių taikymo sričių Lietuvoje yra paminėtos tinklalapyje **InfoBitas** [24]:

- **Kalbinis dialogas su kompiuteriu.** Tai tiesioginis bendravimas su kompiuteriu balsu. Programose yra paruošiami komandų sąrašai ir, vartotojui ištarus vieną iš jų, kompiuteris, priklausomai nuo atpažintos komandos numerio, atitinkamai reaguoja. Komandos gali būti labai įvairios pvz., „parodyk kiek valandų“, „pasakyk kiek valandų“, „pagrok Lietuvos himną“ ir t.t.
- **Balso įrašų stenografavimas.** Programinė įranga, skirta balso įrašams stenografuoti. Gali būti naudojama interviu, posėdžiuose, susirinkimuose ir kt. Tokiose programose kartu su

tekstiniu redaktoriumi yra paleidžiama foninė garsinio išvedimo programa, kuri valdoma iš teksto redaktoriaus aplinkos su makrofunkcijomis.

- **Balsinė interneto naršyklė.** Naršyti internete panaudojant balsą yra patogiu, nes tam pritaikytos programos gali atlikti daug skirtingų komandų: užkrauti pageidautą interneto svetainę, paleisti naršyklę, uždaryti atverstą internetinio tinklalapio langą ir kt.
- **Balsas namų automatikoje.** Labai patogiu ir praktiška buitinius elektros prietaisus valdyti balso komandomis. Pavyzdžiui, įjungti televizorių balsu, išjungti kambaryje šviesą ir pan. Balsu valdoma programinė įranga yra sudaryta iš personaliniame kompiuteryje veikiančios komandų atpažinimo bei valdiklio programos. Per nuoseklų prievadą iš kompiuterio priimamas atpažintos komandos numeris ir per I2C magistralę atliekamas elektrinių įrenginių įjungimas/išjungimas.
- **Asmens atpažinimas pagal balsą.** Teismo ekspertizėje toks atpažinimas vadinamas asmens identifikavimu. Nepriklauso nuo teksto. Identifikavimui šiuo metu naudojama: sonografinis metodas, fonetinis-akustinis metodas, kombinuotas metodas, automatinis metodas [24].

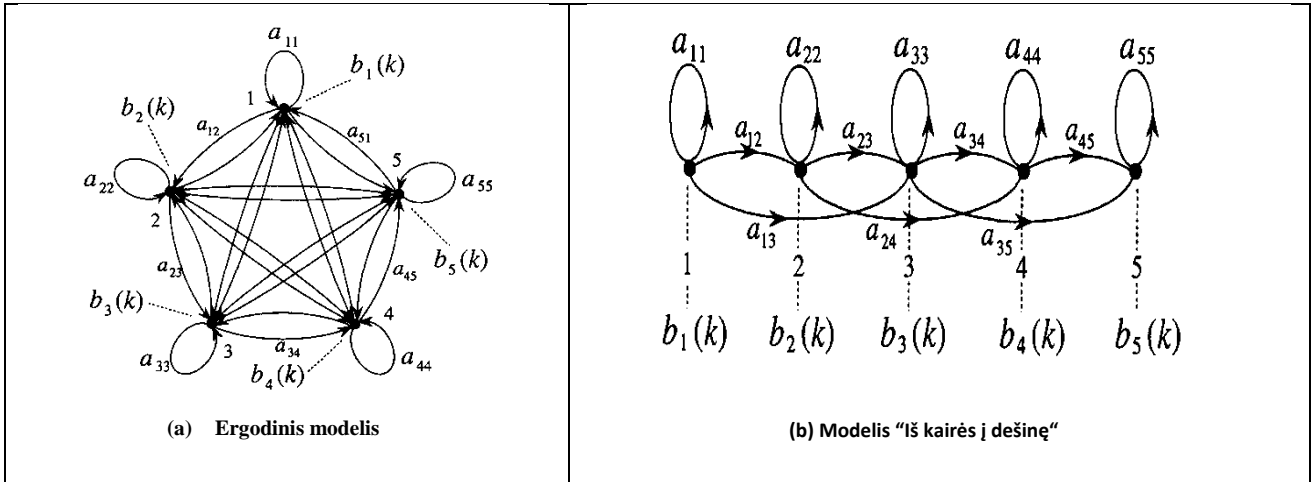
Kadangi automatinio šnekos atpažinimo tyrimų niša niekada nebuvo visapusiškai ištirta, todėl aktyvūs tyrimai nuolatos vykdomi medicinos, farmacijos, multimedijų, telekomunikacijų bei kitose pramonėse ir ypač šiuo metu, kai technologijos yra tobulėjimo viršūnėje. Siekiant visas gyvenimo sritis paversti išmaniomis ir kiekvieną naujai sukurtą prietaisą paversti išmaniuoju „mini žmogumi“, galinčiu kalbėti ir taip pagelbėti žmogui įvairiose gyvenimo situacijose, sukurta gausybė ne tik kalbančių (GPS navigacijos), bet ir galinčių atsakyti (*SIRI* dirbtiniu intelekto mechanizmu pagrįsta programėlė) technologijų. Kalbų įvairovė tokio tipo sistemoms yra tik privalumas.

### 2.3.1. Paslėptieji Markovo modeliai ir jų pritaikymas kalbai atpažinti

Aštuntąjį praėjusio amžiaus dešimtmetį kalbai atpažinti buvo pradėtas taikyti paslėptųjų Markovo modelių metodas. 1975 metais Bakeris aprašė kalbos atpažinimo sistemą *Dragon*, grįstą paslėptaisiais Markovo modeliais, o 1976 m. Jelinekas paslėptųjų Markovo modelių metodą pristatė kaip veiksmingą būdą kalbai atpažinti teigia, Ž. Ringelienė ir M. Filipovič [20], cituodami užsienio mokslininkus (Baker, 1976; Jelinek, 1976). Didžiąją daugumą rinkoje siūlomų automatinio šnekos atpažinimo sistemų dalį sudaro sistemos, besiremiančios paslėptaisiais Markovo modeliais. Tam tikru skaičiumi būsenų perėjimų tikimybėmis gali būti aprašomas fonemų ir alofonų tarimas bei šių tarimų trukmė. Nors dažniausiai PMM naudojami ištisinei kalbai atpažinti, tačiau neretai naudojami ir pavieniams žodžiams. Galima daryti prielaidą, jog tai, kas ir kaip yra tariama, niekada nėra tiksliai žinoma (paslėptas procesas), bet rezultatas visada stebimas (girdimas) ir jį galima užfiksuoti. Taigi automatinio kalbos atpažinimo įtaiso šerdis yra pagal stebėjimo rezultatus sukonstruotas paslėpto proceso modelis [21]. Šis kalbos atpažinimo metodo taikymas tampa vis labiau populiarus, nes matematinis požiūriu jis yra ne tik patogus, bet ir lengvai suprantamas bei įgyvendinamas.

Pastaraisiais dešimtmečiais dažniausiai paslėptųjų Markovo modelių metodas naudojamas kuriant eksperimentines ir komercines sistemas automatinei kalbai atpažinti.

Paslėptųjų Markovo modelių taikymo metodika remiasi tikimybinio modelių sudarymu. Šis atpažinimo metodas yra paremtas statistiniais pavyzdžiais. Šiuo būdu atliekama spektrinė kalbos signalo analizė, kurios metu yra gaunami požymio vektoriai. Vektoriai aprašo įvairius kalbos garsus. Šie garsų modeliai sujungiami į tinklą, atsižvelgiant į jų tvarką įvairiuose žodžiuose.



**2 pav.** Tipinės paslėptųjų Markovo grandinių topologijos

„Paslėptosios Markovo grandinės skirstomos į diskretinius modelius ir tolydinius modelius. Skirstymas atliekamas pagal tai, ar stebimi įvykiai priskiriami kiekvienai būsenai yra diskretiniai ar tolydiniai. Tačiau bet kuriuo atveju pats stebimas įvykis yra atsitiktinis, todėl pats procesas yra dvigubai stochastinis procesas: viena atsitiktinių įvykių seka yra stebima, tačiau antra, vidinė, valdanti stebimą įvykių seką, yra nematoma ir tuo pačiu metu atsitiktinė. Todėl ir pats modelis vadinamas paslėptomis Markovo grandinėmis, nes vidinė mus dominanti procesų seka negali būti stebima tiesiogiai, o yra stebima per atsitiktinį išorinį stebimą procesą. Konkrečiai susiejant tai, kas pasakyta su automatiniu šnekamosios kalbos atpažinimu reikia pasakyti, kad mus domina fonetinių vienetų seka, kuri pati yra atsitiktinis procesas. Tačiau mus dominanti fonetinių vienetų seka negali būti stebima tiesiogiai: atpažinimo sistema fiksuoja požymių vektorius sekas, kurios pačios yra atsitiktinis procesas, ir stebimą atsitiktinį procesą turi susieti su tiesiogiai nestebimais fonetiniais vienetais. Jeigu stebimas požymių vektorius priskiriamas tam tikrai klasei, o po to modeliuojama tokių klasių seka, tuomet paslėpta Markovo grandinė vadinama diskretine, o jei stebimas požymių vektorius iš anksto nėra priskiriamas jokiai klasei, paliekant patį priskyrimą atlikti paslėptai Markovo grandinei.

Diskretinių įvykių PMM yra aprašoma šiais parametrais:

$O = \{O_1, O_2, \dots, O_T, \}$  = stebimų įvykių seka (įėjimo seka).  $T$  – stebimos vektorius sekos dydis (vektorių kiekis).



$Q = \{q_1, q_2, \dots, q_N\}$  = (paslėptų) būsenų skaičius modelyje.  $M$  – klasių skaičius arba vektorinės knygos narių skaičius.

$A = \{a_{ij}\}, a_{ij} = P(q_j \text{ laiko momentu } t+1 | q_i \text{ laiko momentu } t)$  = perėjimų tarp būsenų tikimybių funkcija. Kiekvienam ergodiniam modeliui teisinga  $a_{ij} > 0$  visoms būsenoms  $i, j$ . Kiekvienam iš kairės į dešinę nukreiptam modeliui teisinga  $a_{ij} > 0$ , jei  $i < j$ .

$B = \{b_j(k)\}, b_j(k) = P(v_k \text{ laiko momentu } t | q_j \text{ laiko momentu } t)$  = išėjimo simbolio (klasės) tikimybių pasiskirstymas būsenoje  $j$ .

$\pi = \{\pi_j\}, \pi_i = P(q_i \text{ laiko momentu } t = 1)$  = pradinis būsenų tikimybių pasiskirstymas.

Dažnai PMM aprašyti naudojama kompaktinė užrašymo forma  $\lambda = \{A, B, \pi\}$ . PMM aprašymas be to reikalauja apibrėžti grandinę sudarančių būsenų skaičių, galimų klasių arba simbolių skaičių.  $\lambda$  reikalauja apibrėžti trijų parametrų  $A, B$  ir  $\pi$  tikimybinio tankio pasiskirstymus. Šis parametrų rinkinys įvertinamas iš apmokymo duomenų rinkinio ir apibrėžia tikimybinį matą  $O = (O_1 O_2 \dots O_T)$ , t.y. tikimybę  $P(O/\lambda)$ , kur kiekvienas stebėjimas  $O_t$  yra viena iš galimų klasių arba vektorinio kvantavimo knygos  $V$  simbolis“ [17].

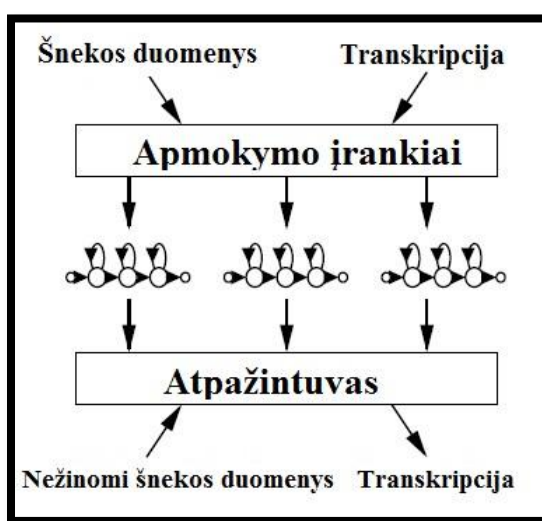
„PMM tipo algoritmai nebūtų tapę tokie populiarūs ir efektyvūs, jeigu nebūtų sukurtas efektyvus apmokymo algoritmas – Baum–Velčo algoritmas [Baum, 1972]. Atpažinimo metu PMM pateikiamas nežinomas ištarimas, kuriam apskaičiuojama tikimybė (panašumas), kad duotoji PMM grandinė sugeneruos tokią stebėjimų seką, kuri atitinka pateiktą nežinomą ištarimą. Tokios tikimybės apskaičiuojamos visoms PMM grandinėms ir išrenkama to ištarimo grandinė, kuriai apskaičiuota tikimybė yra didžiausia. Jeigu atpažinimo sistemos pagrindas paremtas mažesniais negu žodis vienetais, tuomet parenkama PMM grandinių jungtis, kuriai apskaičiuojama suminė didžiausia tikimybė. Labiausiai tikėtinos grandinės išrinkimui naudojamas Viterbi algoritmas: parenkama pora modelio parametrų  $m^*$  ir būsenų sekos  $q^*$  ( $m^*, q^*$ ), kuris tenkina sąlygą:

$$(m^*, q^*) = \arg \max_{(m, q)} P(O, q | \lambda_m) \quad (1)$$

kur (1)  $\lambda_m$  yra  $m$ -toji PMM grandinė ( $m=1, 2, \dots, M$ ),  $M$  – žodyno dydis,  $O = O_1 O_2 \dots O_T$  yra nežinomą ištarimą atitinkanti stebėjimų seka,  $T$  – nežinomo ištarimo pjūvių skaičius,  $q$  – būsenų seka [Viterbi, 1967]. Pati tikimybė  $P(O, q/\lambda_m)$  gali būti apskaičiuota naudojant taip vadinamą judėjimo pirmyn-atgal algoritmą (angl. forward-backward algorithm)“ [17]. „Sintaksė modeliuojama N-gramatikomis, kuriose yra sukaupiamos N paeiliui einančių žodžių statistikos. Atpažinimo procesas grindžiamas Viterbi algoritmu, kai su ištarta fraze dinaminio programavimo būdu lyginami žinomi CD HMM modeliai, surandant panašiausią“ [21]. Kitaip tariant, Viterbi algoritmas gali surasti labiausiai tikėtinus paslėptuose Markovo modeliuose ir išrinkti PMM, turinčią didžiausią tikimybę.

Taigi kalbos atpažinimo metu yra daroma prielaida, kad nežinomas ištarimas atitiks tiriamų požymių vektorių. Seka yra gaunama-generuojama PMM sekos. Skaičiuojamos tikimybės to ištarimo atitikimo akustinių modelių deriniams ir labiausiai tikėtina modelių kombinacija identifikuoja ištarimą [10].

HTK atpažinimo programų rinkinys yra vienas populiariausių ir plačiausiai naudojamų atpažinimo įrankių. HTK yra skirtas kalbos apdorojimo priemonėms, paremtoms PMM, realizuoti. Yra du pagrindiniai apdorojimo etapai: pirmiausiai, HTK mokymo įrankiai yra naudojami siekiant apytikriai apskaičiuoti paslėptųjų Markovo modelių rinkinio parametrus, naudojant mokymui pateiktus pasakymus (įrašus) ir su jais susijusias transkripcijas; antra, nežinomi pasakymai/įrašai yra transkribuojami naudojant HTK atpažinimo įrankius [6]. Žemiau pateikiama schema, vizualiai apibūdinanti šiuos du svarbiausius apdorojimo etapus.



3 pav. HTK atpažinimo įrankio šnekos apdorojimo schema

Atliekant tyrimą su HTK paketu, kiekvienam etapui atlikti reikia panaudoti po atskirą programą iš HTK atpažinimo programų rinkinio. Visos programos yra valdomos iš komandinės eilutės per papildomus failus ir eilutės komandas. Pagrindiniai darbo etapai su HTK programų paketu aprašomi pateikiant pagrindinius darbo etapus su paaiškinimais bei jiems įgyvendinti naudojamų programų pavadinimais:

1. **Žodyno sudarymas.** Kad projektuojama sistema žinotų, kokios žodžių (akustinių vienetų) kombinacijos yra leidžiamos ir kokios – ne;
2. **Požymių apskaičiavimas.** Atpažinimo sistemose naudojami ne patys įrašai, o iš jų apskaičiuoti požymiai, vadinasi, norint apskaičiuoti požymius pirmiausiai reikia turėti įrašų failus. Požymių skaičiavimui naudojama programa *Hcopy*;
3. **Modelių apmokymas.** Apmokymas vyksta naudojant Baum–Welch algoritmą, paleidus programą pavadinimu *HRest*. Svarbus apmokymo etapas – modelių failų kūrimas, kuriuose nurodomi pagrindiniai PMM modelio parametrai. Modelių failai – tai specialaus formato

failai, kuriuose bus surašomi atitinkamo modelio parametrai: perėjimų tikimybės ir išėjimų tikimybės. Kadangi HTK operuoja tik tolydiniais PMM modeliais, tai ir išėjimų tikimybės modeliuojamos Gauso skirstiniais. Vadinasi, išėjimų tikimybės aprašomos vidurkiais ir standartiniais nuokrypiais;

4. **Atpažinimas.** Atlikus apmokymą ir turint testavimui skirtus įrašus galima atlikti testavimą ir patikrinti, kaip gerai vykdomas atpažinimas. Atpažinimui patartina naudoti kitus įrašus, negu buvo naudoti apmokymui. Testavimas vykdomas paleidus programą *HVite*, kuri grindžiama Viterbi algoritmu.

Balso komandų atpažinimas, naudojant paslėptas Markovo grandines, gali būti organizuotas keliais būdais:

- Pritaikant kiekvienam garsui atskirą Markovo grandinę, vėliau ieškant labiausiai tikėtinos šių grandinių sekos. Labiausiai tikėtina grandinių seka turėtų atitikti pasakytą balso komandą sudarančius akustinius vienetus. Šis būdas yra labiau tikėtinas rišlios šnekos ir didelio žodyno atpažinimo atveju. Kiekvieno žodžio modeliavimas atskira grandine yra gana sunkiai įmanomas uždavinys. Dar sudėtingiau yra taikyti modeliavimą viena grandine rišlios šnekos uždaviniuose, nes grandinė turėtų apimti mažiausiai kelių ar net keliolikos žodžių junginius, kurie yra pasakomi be pauzių. Modelių kiekis tampa nesuvaldomu, nes tokių junginių įvairovė yra labai didelė, todėl dažniausiai modeliuojami vieno bei trijų garsų junginiai (tokių modelių kiekis yra ne tik santykinai ribotas, bet ir leidžiantis sumažinti paieškos erdvę).
- Antrasis būdas orientuotas į atskirų balso komandų atpažinimą, kai balso komandų skaičius yra ribotas. Tokiu atveju kiekviena balso komanda atitinkamai modeliuojama viena paslėpta Markovo grandine. Tokio tipo modeliavimas turi savų pranašumų.
- Geriau sumodeliuojami koartikuliacijos efektai: apmokant vientisą Markovo grandinę ją sudarančios būsenos tiksliau modeliuos įvykių dinamiką ir jų tarpusavio sąryšius, lyginant su atveju, kai jungiamos atskiros kelių būsenų Markovo grandinės, atskirai modeliuojančios tam tikrą fonetinį vienetą arba jų junginį.
- Taip pat geriau modeliuojamos pauzės ir kiti akustiniai įvykiai balso komandose, kurias sudaro daugiau nei vienas žodis: balso komandą, turinčią daugiau negu vieną žodį arba žodį, su jame esančiais sprogstamaisiais garsais (atitinkamais sprogstamaisiais priebalsiais), turinčiais nedideles pauzes, galima tikėtis, jog vientisa Markovo grandinė tokius akustinius reiškinius sumodeliuos geriau, nei kelios sujungtos grandinės [19].

Taigi kalbos atpažinimui nepakanka kalbos atpažinimo tikslumą vertinti tik iš to, kaip rezultatus sugeneruoja apmokytas ir neapmokytas profiliai. Todėl anksčiau paminėtos prielaidos apie paslėptąsias Markovo grandines paskatino atlikti eilę eksperimentų naudojant HTK paketą ir įvertinti, ar galima pasiekti geresnius pasirinkto garsyno atpažinimo tikslumo rezultatus modeliuojant

visą balso komandą viena ištisine paslėpta Markovo grandine. Taip pat hibridinio modelio (ispanų kalbos atpažintuvas + HTK paketas) pritaikymas buvo veiksmingas atpažįstant ligų, nusiskundimų ir vaistų pavadinimus *Infobalsas* projekte, todėl daroma prielaida, kad geri rezultatai gali būti gauti tiriant įvairių žodžių garsyną.

### **2.3.2. Garsyno reikšmė automatiniame kalbos atpažinime**

Pasaulyje kalbos atpažinimo technologijų raida jau išibėgėjusi, ji yra aprašyta, nauji ir svarbūs pastarojo meto pasiekimai aprašinėjami, o balso paslaugų sistemų kartų apžvalga ir charakteristikos – pateiktos. Svarbiu faktoriumi kalbos atpažinime išlieka tyrimų objektu laikomi konkretūs įrašai, kurie sudaro garsyną. Garsyną galima apibrėžti kaip specialiai konkrečiam tikslui surinktų kalbos signalų duomenų bazę (angl. speech corpora).

Universalaus garsyno sukūrimas yra sudėtingas uždavinys. Toks garsynas turi gerai aprašyti kalboje sutinkamų fonetinių vienetų įvairovę, diktorių ir kalbėjimo stilių įvairovę, įvertinti kontekstinius efektus ir pan. Dėl šios priežasties garsynai sudaromi orientuojantis į tam tikros uždavinių klasės sprendimą [21]. Savaiame aišku, jog turint garsyną yra svarbu nuodugniai patikrinti ir užtikrinti visų įrašų nepriekaištingą kokybę, nes kalbai atpažinti naudojant tam tikrus garsynus didelę įtaką daro juose esantys triukšmai bei kiti trukdžiai. Gali būti skiriami dviejų tipų garsynai: garsynai, kuriuose surinkti įrašai yra „perskaitytos šnekos“ (nespontaniškos šnekos įrašai) (pvz.: knygų ištraukos, skaičių sekos ir kt.) ir garsynai, kurių turinį sudaro spontaniškos kalbos įrašai (pvz.: dialogai ir kt.). Nespontaniškos šnekos garsynas buvo naudojamas ir šiame darbe tyrimui atlikti. Jis bus aprašytas šiek tiek vėliau.

Pirmasis žinomas garsynas buvo JAV sukurtas TI-DIGITS – izoliuotai tariamų skaitmenų ir jų sekų garsynas. Nuo to laiko sukurta daug garsynų, besiskiriančių savo dydžiais, paskirtimi, anotacijos detalumu. Telefonijos poreikiai skatino kurti įvairius skaitmenų, jų sekų, komandų ir jų sekų garsynus; sukurti mobiliųjų ir stacionariųjų telefonų pokalbių garsynai įvairiose aplinkose – tyloje, biuro darbo, stovinčio ir važiuojančio automobilio aplinkose; kaupiami diktorių įkalbėti specializuoti tekstai, radijo ir televizijos transliacijų garsynai. Nemažai vietos garsynų saugyklose užima specializuoti garsynai, skirti atskleisti tam tikrus šnekos ypatumus: bendrinės šnekos ir tarmių; vyrų, moterų, vaikų; laisvo pokalbio ir skaitomo teksto; pavienių žodžių, komandų ir rišlaus teksto. Universalūs garsynai skirti atskleisti visus bendruosius šnekos ypatumus. Garsynai gali būti anotuoti sakiniiais, žodžiais, skiemenimis, garsais. Galima pastebėti tokias garsynų kūrimo tendencijas: 1) taikomosios paskirties garsynų apimtys nuolat didėja, jų kūrimui dideles investicijas skiria telefonijos, automobilių pramonės firmos, jungiamos mokslo ir verslo pajėgos; 2) vis gausėja nacionalinių garsynų, dokumentuojančių kalbą, kaip tautos socialinės kultūrinės aplinkos paminklą; 3) kuriami gerai anotuoti sisteminiai garsynai, skiriami išsamiems šnekos mokslo tyrimo darbams.

Lietuvoje jau kelis dešimtmečius vykdamat šnekos atpažinimo mokslo tyrimo darbus, dažniausiai būdavo naudojamos nedidelės, konkrečiam uždaviniui spręsti skirtos darbinės garsų įrašų duomenų bazės. Literatūroje yra aprašyta tik viena lietuviška skaitmenų ir komandų tarimo duomenų bazė LTDIGITS, palygintina su amerikietiškuoju TI-DIGITS garsynu. Garsynas sukonstruotas taip, kad atspindėtų visus svarbiausius bendrinės lietuvių kalbos fonetinius ypatumus skirtus: lietuvių kalbos informacinių technologijų mokslo tyrimui ir taikomiesiems darbams. Garsynas aprūpintas originalia garsų paieškos programine įranga Tescal [18].

Susipažinus su B. Balvočiaus ir L. Telksnio medžiaga apie garsynus [1], paaiškėjo, jog TIMIT (ir jo variantai NTIMIT, CTIMIT) šnekos duomenų bazė – garsynas – laikomas klasikinio garsyno pavyzdžiu. TIMIT garsynas yra hierarchinės duomenų bazės pavyzdys, kur svarbų vaidmenį atlieka failų sistemos hierarchinė struktūra ir ten sudėta informacija. Duomenų bazės tikslas – pateikti duomenis kalbos atpažinimo sistemų kūrėjams, fonetikams ir kitiems tyrinėtojams. Įrašomos ištaisai sakomos frazės, kurias diktorius skaito. Duomenys yra atskiruose failuose, kuriuos galima rasti atskirose direktorijose ir visa duomenų bazė yra sukelta į CD-ROM diskus.

Aišku yra tai, jog garsynų įvairovė yra didelė. Įvairios organizacijos platina jau sukurtus garsynus, dar yra daugybė, kurių platinimu rūpinasi patys garsynų sudarytojai. Dažnai garsynai gali būti ir neplatinami, tarkime, tokie, kuriuos asmeniniams tyrimo tikslams sukuria studentai. Kiekvienas garsynas yra savotiškas, turintis jam būdingas charakteristikas, tačiau B. Balvočius ir L. Telksnys [1] apibrėžia reikalavimus, pagal kuriuos nusakoma, koks garsynas galėtų būti efektyviai naudojamas šnekos atpažinimo sistemoms kurti. Taigi garsynas turėtų tenkinti šias sąlygas:

- suteikti vartotojui galimybes manipuluoti dideliais duomenų kiekiais (perkelti, atnaujinti, naikinti) ir užtikrinti duomenų korektiškumą bei integralumą;
- turėti efektyvią paieškos sistemą, užtikrinančią tiek šnekos signalų, tiek meta-duomenų ištraukimą;
- turėti savybes, atitinkančias bendro darbo sistemos principus (daugivartotojiškumas, duomenų kontrolė ir dalinimasis);
- turėti standartinę sąsają duomenų lygyje (duomenims importuoti ir eksportuoti į kitas sistemas); Sistemos atvirumas nulemia duomenų pasikeitimo su kitomis sistemomis galimybes.
- turėti standartinę programinę sąsają, kad, sukūrus papildomą funkcionalumą, jis taptų lengvai prieinamas ir kitiems vartotojams;
- turėti vartotojo sąsają patogiam duomenų bazės užklaudemui ir ataskaitų formavimui. Žmogus, tiriantis šnekos signalus, nebūtinai turi mokėti užklaudem kalbą, kad susirastų ir peržiūrėtų informaciją ar ją paruoštų eksperimentui. Patogi vartotojo sąsaja leidžia koncentruotis į tyrimą, o ne į duomenų tyrimui paruošimą [1].

Kadangi nėra garsyno, kuris tenkintų visus reikalavimus, todėl verta siūlyti naujas duomenų modelių ir programinės įrangos architektūrų kombinacijas. Gero garsyno pritaikymas, atliekant šnekos atpažinimo tyrimus leidžia pakankamai patikimai įvertinti daugiadiktorinio lietuviškų balso komandų atpažinimo tikslumą skirtingose situacijose.

### 2.3.3. Teksto transkribavimas

Teksto transkribavimas – tai procesas, kurio metu tekstas yra pakeičiamas fonetinių vienetų seka. Transkripcija yra reikalinga, nes žodžius sudarančios raidės dažnai nenusako tiesioginio perėjimo prie garsų (rašybos taisyklės paprastai būna daug konservatyvesnės (lėčiau kintančios) nei šnekamoji kalba). Lietuvių kalba yra unikali, nes ji nesutampa su jokios kitos kalbos transkripcijomis [21].

Kaip teigia Kasparaitis, kalbėdamas apie transkribavimą [8], galima išskirti du pagrindinius transkribavimo metodus: pagrįstą žodynu ir pagrįstą taisyklėmis. Lietuvių kalbai geriau taikyti formaliomis taisyklėmis paremtą metodą, užuot taikius žodynu paremtą metodą. Tam paaiškinimas yra paprastas – lietuvių kalbos žodžiai yra kaitomi ir lietuvių kalbos transkripcija yra nesudėtinga, jeigu žodžius galima laikyti suskiemenuotais ir sukirčiuotais. Žodynu pagrįstas metodas atitinkamai yra pritaikomas kalbose, kuriose žodžiai turi nedaug formų ir tos pačios raidės gali atitikti skirtingus garsus (pvz., anglų) [21].

Pasak Kazlauskienės ir kt., lietuvių kalbos rašyba remiasi šiais principais:

- fonetiniu (arba fonologiniu): rašoma yra taip, kaip tariama;
- morfologiniu (arba morfeminiu): morfemas (mažiausias reikšminės žodžio dalis) stengiamasi užrašyti vienodai, nors jos gali būti skirtingai tariamos;
- tradiciniu: kai kurių rašmenų vartojimas ar nevartojimas paaiškinamas esant tradicija, jog įprasta taip rašyti [9].

Transkripcijų nauda kuriant garsynus yra akivaizdi. Generuojant garsynus, visą su garsynais susijusią informaciją reikia koduoti universaliu būdu, kad pagamintas produktas būtų suprantamas bet kurios kitos kalbos garsynų kūrėjams. Garsynas turi būti universalaus pritaikymo ir todėl, kad vartotojams labai svarbu turėti galimybę palyginti įvairių kalbų fonetines sistemas bei pritaikyti savo darbe kitų kalbų tyrimui sukurtą programinę įrangą. Dėl šių priežasčių atsiranda vis daugiau bandymų sudaryti universalias fonetinės transkripcijos sistemas, kurios bendrais visiems kompiuteriams ASCII kodais koduoja kiekvienos kalbos specifines tarties ypatybes [18].

Be ASCII kodavimo sistemos universaliai naudojimui gali būti taikomas vienas iš tikslingiausių transkripcijų sudarymo metodų naudojantis UPS (Universal Phone Set) simbolių rinkiniu. UPS yra universalių fonemų aibė, kurią lengvai supranta kompiuterinės sistemos. Šis fonemų rinkinys yra sudarytas IPA pagrindu, papildytas SAPI garsais ir žymomis, kurių nėra IPA fonemų sąraše. UPS gramatikose žodžiai gali būti papildomi įvairiomis žymomis nurodant kirtį,

tarimą ir kitas savybes [13]. Anksčiau UPS transkripcijas buvo galima sudaryti rankiniu būdu, dabar naudojantis Microsoft Speech Server su MS Visual Studio programiniu paketu, kuriame automatiškai sugeneruojamos UPS transkripcijos pagal integruotus UPS fonemų sąrašus. Teisingas transkripcijų parinkimas/sugeneravimas gali ženkliai pagerinti kalbos atpažinimo tikslumą [3].

Tyrimams su garsynais panaudojant ispanų kalbos atpažintuvą, kiekvienai balso komandai taip pat bus sukuriama transkripcijos. Jos bus sukurtos ir toms gramatikoms, kurios bus panaudotos „Demo“ programoje.

#### **2.4. Analitinės dalies išvados**

1. Užsienio kalbų mokymo procese svarbu užtikrinti ne tik kalbinės, bet ir kultūrinės komunikacijos gebėjimus, nes jie, mokantis kalbos, yra vienas nuo kito neatskiriami.
2. Galimybė užsienio studentams mokytis kalbos naudojantis „SurfaceLanguages“ ir „Loecsen“ priemonėmis yra reali, nes jos yra nemokamos ir be apribojimų prieinamos internete. Priemonių trūkumas – pasigendama į tarimo įgūdžius orientuotų užduočių, nes priemonės neturi savyje integruoto kalbos atpažinimo įrankio.
3. Kalbos mokymo priemonėje „Duolingo“ derinami klausymo, vertimo, kalbėjimo žodžiu ir kiti pratimai, leidžiantys visapusiškai lavinti pasirinktos kalbos žinias. Priemonės trūkumas – joje nėra galimybės mokytis lietuvių kalbos. Rekomenduotina kuriant virtualų šnekamosios lietuvių kalbos kursą remtis „Duolingo“ priemonėje pateiktomis idėjomis ir įterpti užduočių, turinčių kalbos atpažinimo galimybę.
4. „Oneness“ yra virtualus kalbų mokymosi kursas, kuriame besimokantysis gali įgyti lietuvių kalbos žinių. Kurse pateikiama medžiaga visoms besimokančiojo kalbinėms kompetencijoms lavinti, bet kalbiniam įgūdžiam lavinti nenaudojamos kalbos atpažinimo priemonės. Siekiant išvengti medžiagos dubliavimosi, kuriamame kurse būtų galima kelti tiesiogines nuorodas į „Oneness“ kurso medžiagą.
5. Kalbos atpažinimo pritaikymo sričių įvairovė plečiasi, todėl specialistai yra suinteresuoti kalbos atpažinimo tyrimais bei jų pritaikymu praktikoje. Automatinis kalbos atpažinimas – pirmoji balsinių technologijų produktų grandis, įgalinanti technologijas suprasti žmogaus kalbą. Viena iš kalbos atpažinimo panaudojimo sričių yra kalbų mokymas, lavinant kalbos tarties įgūdžius.

### 3. TYRIMŲ DALIS

Kaip jau buvo minėta įvade, esant užsienio studentų poreikiui mokytis lietuvių kalbos, šią problemą galima išspręsti naudojantis įvairiomis technologijomis, specialiai pritaikytomis prie inovatyviai pateiktos mokomosios medžiagos, padėsiančios nesunkiai įgyti pradinių lietuvių kalbos žinių (bendram kalbos suvokimui ir susikalbėjimui).

Siekiant kuo aukštesnių virtualiu būdu įgytų kalbos rezultatų, svarbu ugdyti visas kalbines kompetencijas. Dėl šios priežasties, virtualiame kurse svarbu pateikti medžiagą interaktyviai, o kalbines kompetencijas ugdyti pasitelkus asmens kalbos atpažinimą. Šiam tikslui įgyvendinti svarbu sukurti garsynus ir su jais atlikti tyrimus, kurių rezultatai nukreips reikiama kryptimi, kuriant kalbos įgūdžiams formuoti skirtas užduotis, vėliau jas tikslingai pritaikant virtualiame kurse.

Šiame skyriuje aprašomi vardų ir pavardžių atpažinimo tyrimai, naudojant ispanišką apmokytą profilį bei HTK programų rinkinį. Tyrimuose remiamasi 2014 m. bakalauro darbe atliktų tyrimų nustatytais parametrais. Pateikiama tyrimo eiga ir metodika, tyrimų analizės ir rezultatai. Sprendžiami ir įgyvendinami įvade iškelti darbo uždaviniai.

#### 3.1. Tyrimo metodologija

Visi darbe atliekami tyrimai yra paremti atitinkama tyrimo metodologija, kuri aiškina darbui atlikti naudojamus metodus ir techniką. Šio darbo tikslui įgyvendinti bei įvade iškeltiems uždaviniams atlikti darbe bus naudojami tokie tyrimo metodai:

- deskriptyvinis
- eksperimentinis
- kiekybinės analizės

**Deskriptyvinis metodas.** Deskriptyvinio (aprašomojo) tyrimo esmė – surinkti informaciją apie esamą situaciją, taip pat įvertinti situaciją dominančių požymių atžvilgiu. Šiuo atveju deskriptyvinis metodas taikomas įvairioms teorijoms aiškinti, analogiškų tyrimų apžvalgai, svarbiausiems terminams apibrėžti, vardų atpažinimo galimybėms ir gautiems rezultatams įvertinti. **Eksperimentinis metodas** – reikalingas surinktos medžiagos tyrimams, rezultatų priežasčiai nustatyti, keliamoms hipotezėms pagrįsti ir tolimesnių tyrimų idėjoms bei pakraipoms, siekiant gauti pageidautiną tyrimų baigtį. **Kiekybinės analizės metodas** – tai struktūrizuotas, besiremiantis iš mokslinės problemos kylančia hipoteze tyrimas, taikant matematinius analizės metodus tyrimo duomenims apdoroti bei nagrinėjamam reiškiniui aprašyti.

Remiantis nusakytais metodais atliekamas įdiktuočių pavardžių ir vardų bei tų pačių iškraipytų pavardžių ir vardų atpažinimo tikslumo tyrimas. Atliekamiems tyrimams naudojamas garsynas, kuris sudarytas iš 10 skirtingų taisyklingai išstartų pasirinktų pavardžių ir vardų balso komandų. Kiekviena pavardė ir vardas įdiktuojami po 20 kartų. Kadangi garsyno kūrime dalyvavo 10 diktorių, darbui naudojamų įdiktuočių pavardžių ir vardų įrašų iš viso yra 5200. Tolimesniems tyrimams naudojamas



dar vienas sukurtas garsynas, kuris taip pat sudarytas iš 10 skirtingų pavardžių ir vardų balso komandų, tačiau šiuo atveju jos ištartos netaisyklingai. Kiekviena pavardė ir vardas įdiktuojami po 10 kartų. Šio garsyno kūrimo dalyvavo 5 diktoriai, naudojamos balso komandos buvo įdiktuos 500 kartų.

Antrasis etapas ruošiantis tyrimui yra gramatikų ir transkripcijų kūrimas, kuris siejamas įdiktuoju garsynu ir jo atpažinimo gerinimu darbo eigoje. Tam paruošiama žodinė gramatika bei gramatika, perrašyta transkripcijomis.

Tyrimai vykdomi vienodomis sąlygomis, testavimai atliekami naudojant ispanų kalbos balso atpažintuvą, apmokytą anglų kalba. Norint atlikti profilio apmokymą diktuojant frazes ispanų kalba, reikalinga, kad jos būtų diktuojamos ispanakalbio. Tyrimui išbandomas savu balsu apmokytas profilis. Rezultatams, išgautiems su ispanų kalbos atpažintuvu, palyginti, įsisavinamas HTK atpažinimo programų rinkinys ir su juo atliekamas dalies garsyno testavimas.

Dėl didelio kiekio duomenų parengtos specialios programos, t.y. tyrimas *batch* failais (plėtinys *.bat*), apdorojančios duomenis ir pateikiančios rezultatų sąrašus tekstiniuose failuose, kuriems reikalinga statistinė analizė. Tokių failų paskirtis yra automatizuoti kokį nors procesą, tačiau jie nėra labai palankūs dirbant su dideliais kiekiais duomenų, nes pačių rezultatų analizė ir sisteminimas rankiniu būdu užtrunka labai ilgai. Šių programų privalumas – nereikia testuoti kiekvieno garsinio failo atskirai, komandų nustatymuose nurodomas takas, iš kur ir kokie duomenys turi būti paimiti. Tai paspartina garsyno testavimo eigą.

**Demonstraciniai rezultatai.** Įgyvendinta idėja sukurti programą, kaip, remiantis atliktais tyrimų rezultatais, veikia įvairių lietuviškų žodžių komandų atpažinimas, kai ištariama komanda iš nurodyto sąrašo laukiant programos atsako į komandą. Tikslaus atpažinimo atveju, programa turi išvesti ekrane įdiktuoją komandą ir atpažinimo tikslumą. Programa naudojama lietuvių kalbos mokymo kurse, ją įkomponuojant prie atitinkamų kurso temų, tikslingai demonstracijoje pritaikant gramatikas, kurios sudarytos iš skirtingos leksikos ir atitinka mokymo temą.

### **3.2. Garsynų ir gramatikų paruošimas pavardžių ir vardų atpažinimo tyrimams**

Siekiant sukurti kursą, kuriame būtų įkomponuota lietuvių kalbos atpažinimo funkcija, visų pirma reikalinga atlikti tyrimus su sudėtingomis lietuviškomis pavardėmis ir vardais, norint įsitikinti, kad kalbos atpažinimas kurse bus veiksmingas atpažįstant nesudėtingus lietuviškus žodžius. Yra žinomas faktas, kad užsieniečiams lietuvių kalba yra sudėtinga: sunku suprasti kalbos gramatiką, taisyklingai ištarti žodžius ir pan. Dar sunkiau nei ištarti žodžius yra ištarti lietuviškus pavardes ir vardus. Dėl šios priežasties kalbos atpažinimo tyrimai vykdomi su lietuviškomis pavardėmis ir vardais, kurie yra sunkiai ištariami, nes susideda iš dvibalsių ir diakritinius ženklus turinčių raidžių, ir su iškraipytais tais pačiais pavardėmis ir vardais, simuliuojant užsieniečių vardų ir pavardžių tarimą.

Vykdomi lietuviškų pavardžių ir vardų atpažinimo tyrimai būtų neįmanomi be daugiadiktorinių pavardžių ir vardų garsynų. Taisyklingai išstartų vardų ir pavardžių garsyno sudaryme dalyvavo 10 diktorių (KTU bendruomenės nariai), t.y. 6 moteriški balsai ir 4 vyriški. Balso komandų įrašinėjimas vyko naudojant programą, veikiančią MS DOS operacinės sistemos aplinkoje.

Garsynas yra suskirstytas į katalogus pagal diktorius. Kiekviename tokiam atskiro diktoriaus kataloge yra 10 skirtingų atskirų pavardžių ir vardų katalogai, kuriuose galima rasti:

- audio failo turinį aprašantį pagalbinį failą „mano txt“, kuriame nurodytas tekstas apie išstarto vardo kiekį;
- audio failas „mano voc“ (16000 Hz, 16 bitų mono audio formatas) ir tas pats failas WAV formate „mano wav“;
- anotacijų failas „mano zgl“, kuriame pateikti kiekvieno įrašo pradžios ir pabaigos adresai;
- pagalbiniai failai, naudoti garsyno rinkimo ir tikrinimo metu.

Tam tikro diktoriaus įrašus identifikuojantis katalogo pavadinimas yra sudarytas iš diktoriaus lyties pirmosios raidės (M – vyras (male), F – moteris (female)) bei vardo ir pavardės pirmųjų 3 raidžių, pvz., FIVEVAL arba MKASRAT. Vardus talpinantys katalogai pavadinti kataloge sudėtų ištarimų vardu, tarkime, jeigu kataloge ištarimai yra „adomaviciute inga“, tai katalogas pavadintas „ADOMAVICIUTE“ ir pan. Garsyną sudarantis katalogų sąrašas bei jame esantys ištarimai pateikiami 1 priede.

Iš viso yra 10 skirtingų balso komandų. Sudarant garsyną pasirenkama kataloge esanti įrašinėjimo komanda ir pradedama diktuoti komanda. Kiekvieną balso komandą reikia diktuoti 20 kartų, pradedant skaičiuoti nuo 0 iki 19 kartų. Diktuojant įrašus labai svarbu kuo taisyklingiau ir aiškiau išstarti kiekvieną įrašą, pabrėžiant tariamo įrašo pradžią ir pabaigą, kad būtų užtikrintas tikslus ir kokybiškas atpažinimo tyrimas. Įrašius įrašus galima iš karto juos perklausyti ir pataisyti tuos, kuriuose netiksliai išstarta komanda arba girdimas pašalinis triukšmas. Tam atlikti nereikia perdiktuoti visų įrašų, nes kiekviename kataloge yra programėlė, leidžianti perrašyti blogą įrašą, nurodant jo numerį.

Kalbos atpažinimo tyrimai būtų neįmanomi be tyrimus papildančio dar vieno garsyno, kurį sudaro tokios pačios balso komandos kaip ir pirmajame garsyne, tačiau šiuo atveju jos išstartos labai netaisyklingai, simuliuojant užsienio studento ištarimą. Šio tipo garsyno sudaryme dalyvavo 5 diktoriai: 2 vyriški balsai, likusieji – moteriški, o kiekviena balso komanda įdiktuota po 10 kartų. Kadangi šis garsynas buvo įdiktuotas asmenų iš artimos asmeninės aplinkos, balso komandų įrašinėjimas vyko kiekvienam individualiai naudojant programą „Wavesurfer“. Diktuojant balso komandas prisilaikyta to paties principo kaip ir pirmuoju atveju, t.y. garsynas suskirstytas į katalogus pagal diktorius. Tačiau, skirtingai nei pirmuoju atveju, čia kiekviename tokiam atskiro diktoriaus

kataloge yra sukeltos visos 10 skirtingų komandų. Tokiuose kataloguose galima rasti tik audio failą (16000 Hz, 16 bitų mono audio formatas) išsaugotą WAV formatu.

Atliekant kalbos atpažinimo tyrimus ne mažiau svarbu atpažinimo sistemai paruošti gramatikos failą. Šis etapas yra svarbus tuo, jog gramatika – tai failas, kuriame nurodomos komandų transkripcijos ir savotiškos funkcijos, nurodančios, ką reikia išvesti į ekraną įvykus atpažinimui. Gramatika sudaroma transkripcijomis, kurios parašytos naudojantis sintezatoriumi, laisvai prieinamu internete, t.y. svetainėje <http://www.acapela-group.com/>. Šio sintezatoriaus veikimas pagrįstas *text-to-speech* principu. Komandoms transkripcijos rašomos intuityviai, vėliau perklausant, kaip jas ištaria sintezatorius. Panašiausi į reikiamą komandą ištarimai yra panaudojami gramatikai užrašyti kaip alternatyvos. Žemiau pateikiama ištrauka iš gramatikos failo:

```
<item><one-of><item>alesandravychute_ramjute</item>  
<item>aleksandraviciute_ramute</item>  
</one-of><tag>$.value = "aleksandraviciute ramute"</tag></item>
```

### 3.3. Pavardžių ir vardų apmokyto profilio „Default\_Iveta1” įtaka atpažinimo tikslumui

Tyrimams naudojamas ispanų kalbos atpažintuvas yra vienas iš įrankių kalbos atpažinimo procese. Veikimo ypatybėms bei profilių įtakos atpažinimo kokybei nusakyti pirmiausia tyrimą reikia pradėti nuo profilių pritaikymo darbui. Vienas iš tyrime reikalingų profilių yra profilis „Default\_Iveta1”. Šis profilis sukuriamas remiantis tokia tvarka: *Control panel* → *Speech recognition* → *Advanced Speech options* → *Speech recognition*. Įvykdžius šią veiksmų seką, atverstame lange pasirenkamas ispanų kalbos atpažintuvas *Microsoft Speech Recognizer 8.0 for Windows (spanish – Spain)* ir sukuriamas naujas profilis: *New* → *profilio pavadinimas (Iveta1)* → *OK*. Tam, kad sukurtas profilis būtų apmokytas (įgautų balsinių savybių), pabaigus profilio kūrimo darbus po funkcijos → *OK* seka funkcija *Train Profile*. Paspaudus *Train Profile* mygtuką atliekamas kompiuterio pateiktų balso komandų diktavimas anglų kalba. Tokiu būdu kompiuteris geba įsisavinti diktuojančiojo balso savybes ir panaudoja jas atpažinime nepaisant to, jog jos yra diktuojamos ne ispanų kalba. Norint atlikti profilio apmokymą diktuojant ispanų kalbos frazes, reikalinga, kad jos būtų diktuojamos ispanakalbio. Atlikus minimalius profilio paruošimo darbus pradedami tyrimai su apmokytu profiliumi.

Pirmiausia tyrimai su apmokytu profiliumi atliekami naudojant taisyklingai išstartas balso komandas. Abiem atvejais, tiek taisyklingai išstartų, tiek klaidingai išstartų balso komandų atveju tikrinama hipotezė, kuriuo atveju: naudojant – HTK programų paketą ar ispanų kalbos atpažintuvą balso komandos atpažįstamos tiksliau. Naudojant apmokytą profilį *Iveta1*, apmokytą *fiveval* diktoriaus balso savybėmis, tyrime naudojama dalis vardų ir pavardžių garsyno, įrašyto moteriškais ir vyriškais balsais, diktorių: *fdeilei*, *fginbar*, *fiveval*, *mdomvas* ir *mkasrat*. Dalies garsyno pakanka, kad būtų patikrinta apmokyto profilio galimybė atpažinti balso komandas, taip pat ir tolimesniuose HTK

tyrimuose bus naudojama tik dalis balso įrašų tyrimams, nes likusieji dalyvaus apmokymo procese. Žemiau esančioje lentelėje pateikiami susisteminti rezultatai, koks atpažinimo tikslumas buvo pasiektas testuojant balso komandas su apmokytu profiliu.

**1 lentelė** Tyrimo su teisingomis balso komandomis rezultatai

Vardas	Atpažinimo tikslumas su apmokytu profiliu
Adomavičiūtė Inga	97%
Aleksandravičiūtė Ramutė	100%
Ališauskienė Jadvyga	100%
Bulachienė Danguolė	92%
Čizaitė Radvilė	96%
Danilkevičius Rytis	99%
Eidžiūnas Aurimas	100%
Grigaravičienė Jūratė	100%
Ignatavičius Gintaras	100%
Ivancevičiūtė Erika	100%
<b>Bendras vidurkis</b>	<b>98,4%</b>

Atlikus tyrimą su apmokytu profiliu gaunamas rezultatas siekia **98,4 proc.** atpažinimo tikslumą. Tai yra palyginti aukštas rezultatas, kas reiškia, jog atpažintuvas palankiai reaguoja į ilgas balso komandas, turinčias dvigarsių, dvibalsių ir diakritinių ženklų. Visos balso komandos buvo atpažintos beveik idealiai. Blogiausiai atpažintos komandos *Bulachienė Danguolė*, *Čizaitė Radvilė*, *Adomavičiūtė Inga* ir *Danilkevičius Rytis*.

Antruoju atveju buvo atliktas tyrimas, kai naudojant apmokytą ispanų kalbos atpažintuvą testuojamos iškraipytos balso komandos.

**2 lentelė** Tyrimo su iškraipytomis balso komandomis rezultatai

Vardas	Atpažinimo tikslumas su apmokytu profiliu
Adomavičiūtė Inga	68%
Aleksandravičiūtė Ramutė	60%
Ališauskienė Jadvyga	58%
Bulachienė Danguolė	60%
Čizaitė Radvilė	56%
Danilkevičius Rytis	58%
Eidžiūnas Aurimas	52%
Grigaravičienė Jūratė	56%
Ignatavičius Gintaras	58%
Ivancevičiūtė Erika	58%
<b>Bendras vidurkis</b>	<b>58,4%</b>

Gautas atpažinimo rezultatas siekia **58,4 proc.** atpažinimo tikslumą. Tai beveik dvigubai prastesnis rezultatas, lyginant jį su pirmuoju tyrimu. Vadinasi, apmokytas ispanų kalbos profilis geriau atpažįsta taisyklingai tariamas balso komandas.

Tyrimų metu surasti gerai ir blogai išstartų komandų atpažinimo panašumo matų vidurkiai. Gerai atpažintų komandų panašumo matas yra 594,638, blogai atpažintų 472,8705. Turint panašumo matus, surastas slenkstis, pagal kurį galima būtų priimti sprendimą, „teisingai“ ar „neteisingai“ atpažinta balso komanda. Slenkstis yra lygus abiejų atpažinimo atvejų vidurkių sumos pusei. Šiuo atveju slenkstis yra 533,754.

Iš rezultatų, gautų su kitakalbiu atpažintuvu analizės galima spręsti, kad gerų ir blogų ištarimų atpažinimo panašumo matai persidengia, vienas nuo kito mažai skiriasi, todėl slenkščio neįmanoma panaudoti.

### **3.4. HTK atpažinimo programų rinkinio praktinis pritaikymas ir įsisavinimas**

HTK kaip įrankis, jo pritaikymas kalbos atpažinime ir pagrindiniai bruožai yra aptarti teorinėje darbo dalyje, tačiau taip pat svarbu ir praktinis įrankio pritaikymas. Taigi šiame skyriuje aptariamas kalbos atpažinimo programų rinkinio HTK praktinis pritaikymas, įsisavinimas ir dalies garsyno testavimas. Tyrimui naudojama 10 diktorių ištarimai. Bandymai, HTK paketo veikimo įsisavinimas ir modelių apmokymas vyksta pagal 8 diktorių įrašus, t.y. *fdeimak*, *fginbar*, *mdombak*, *malgzil*, *frimram*, *fievzil*, *fraival*, *mvilram* (moteriški ir vyriški balsai). Testavimui naudojami 2 diktorių įrašai, t.y. moterišku balsu diktorės *fiveval* įdiktuoti įrašai bei vyrišku balsu įdiktuoti diktoriaus *mksrat* įrašai.

Pradedant darbą su atpažinimo programų rinkiniu HTK, kaip ir bet kokiai atpažinimo sistemai, reikalinga paruošti keletą išankstinių failų. Pirmasis etapas yra gramatikos failo kūrimas, kuriame bus išvardinta, kokie žodžiai ir kokia tvarka gali būti pateikti atpažinimo sistemai. Taigi sukuriamas failas pavadinimu *gram.dict*, kuriame nurodoma 10 komandų, su kuriomis vyks tyrimas. Kadangi šias komandas sudaro ne vienas žodis, failo *gram.dict* turinys atrodo taip:

```
$word=adomaviciute_inga|aleksandraviciute_ramute|alisauskiene_jadvyga|bulachiene_danguol  
e|cizaite_radvile|danilkevicius_rytis|eidziunas_aurimas|grigaraviciene_jurate|ignatavicius_gintaras|iv  
anceviciute_erika;
```

(\$word)

Tokiu užrašymu sistemai nurodoma, kad bus pasakyta pavardė ir vardas ir jie bus vieni iš nurodytųjų, t.y. bus arba adomavičiūtė inga, aleksandravičiūtė ramutė, ališauskienė jadvyga, bulachienė danguolė, čižaitė radvilė, danilkevičius rytis, eidžiūnas aurimas, grigaravičienė jūratė, ignatavičius gintaras arba ivancevičiūtė erika.

Galimas ir toks gramatikos pateikimo variantas, jeigu vieno ištarimo metu norima pasakyti du žodžius iš pateikto sąrašo:

\$word=adomavičiūtė\_inga|aleksandravičiūtė\_ramutė|ališauskienė\_jadvyga|bulachienė\_danguolė|čižaitė\_radvilė|danilkevičius\_rytis|eidžiūnas\_aurimas|grigaravičienė\_jūratė|ignatavičius\_gintaras|ivancevičiūtė\_erika;

\$word \$word

Tokiu atveju turėtų būti sakoma taip: „adomavičiūtė inga, aleksandravičiūtė ramutė“. Šiam tyrimui bus naudojamas paprasčiausias, pirmasis gramatikos variantas, nes siekiama įsisavinti HTK programų rinkinio veikimą ir pritaikyti daliai garsyno testuoti.

Antrasis etapas yra žodžių tinklo wordnet.txt failo sukūrimas, kuriam reikalinga programa *HParse.exe*. Visos komandos vykdomos naudojantis Far Commander programa, kuri suteikia galimybę lanksčiai ir patogiai dirbti su HTK priemonėmis. Taigi komandų eilutėje nurodomos tokios užduotys: *HParse.exe gram.dict wordnet.txt*. Čia *HParse* yra programa, kuri reikalinga žodžių tinklui *wordnet.txt* sudaryti, *gram.dict* yra pirmame etape sudaryta „gramatika“, *wordnet.txt* yra failas, kurį programai reikia sugeneruoti. Paleidus programą *Hparse*, failas užpildomas tam tikromis žymomis.

Trečiasis etapas yra požymių skaičiavimas. Kadangi atpažinimo sistemose naudojami ne patys įrašai, o iš jų apskaičiuoti požymiai, norint apskaičiuoti požymius, reikalinga padaryti script-failus, kuriuose kompiuteriui nurodoma, kur yra įrašai. Taip pat nurodomi požymių failai, kurie vėliau paleidus atitinkamą programą bus sugeneruoti ir papildyti tam tikrais įrašais. Taip atrodo ištrauka iš failo *failai.scf* turinio:

D:\DARBAS\HTK\HTK\irasai\0A00.wav D:\DARBAS\HTK\HTK\irasai\0A00.mfc

D:\DARBAS\HTK\HTK\irasai\0A01.wav D:\DARBAS\HTK\HTK\irasai\0A01.mfc

...

D:\DARBAS\HTK\HTK\irasai\0A16.wav D:\DARBAS\HTK\HTK\irasai\0A16.mfc

D:\DARBAS\HTK\HTK\irasai\0A17.wav D:\DARBAS\HTK\HTK\irasai\0A17.mfc

D:\DARBAS\HTK\HTK\irasai\0A18.wav D:\DARBAS\HTK\HTK\irasai\0A18.mfc

D:\DARBAS\HTK\HTK\irasai\0A19.wav D:\DARBAS\HTK\HTK\irasai\0A19.mfc

Toks užrašymas reiškia, kad bus paimamas atitinkamas įrašas iš nurodytos vietos (*D:\DARBAS\HTK\HTK\irasai*) ir tam tikromis žymomis bus užpildytas failas su to paties įrašo pavadinimu, tik, šiuo atveju, su plėtiniu *.mfc*. Tokiu būdu programa sukurs užpildytus požymių failus visiems nurodytiems įrašams. Požymių skaičiavimui naudojama programa *Hcopy.exe*. Komandinėje eilutėje nurodomos tokios komandos: *Hcopy.exe -C config -S failai.scf*. Čia *Hcopy.exe* yra programa, reikalinga požymiams sugeneruoti, faile *CONFIG* yra pateikta informacija apie požymių skaičiavimo būdus, failas *failai.scf* yra jau anksčiau minėtas script failas, nurodantis įrašų buvimo vietą.

Ketvirtasis etapas yra modelių failų parengimas pagal specialią formą. Jie reikalingi tolimesniam etapui, kuriame bus apmokoma atpažinimo sistema. Kiekvienam žodžiui sukuriamas atskiras modelio failas pavadinimu *hmm\_\**. Taip atrodo šiam tyrimui paruošti modelių failai:

*hmm\_adomaviciute\_inga*

*hmm\_aleksandraviciute\_ramute*

*hmm\_alisauskiene\_jadvyga*

*hmm\_bulachiene\_danguole*

*hmm\_cizaitė\_radvile*

*hmm\_danilkevicius\_rytis*

*hmm\_eidziunas\_aurimas*

*hmm\_grigaraviciene\_jurate*

*hmm\_ignatavicius\_gintaras*

*hmm\_ivanceviciute\_erika*

Kiekviename šių failų reikia nurodyti, kiek būsenų bus naudojama žodžiui modeliuoti, kiek požymių bus naudojama požymių vektoriuje, nurodomi pradiniai vidurkių ir dispersijų vektoriai bei perėjimų tikimybių reikšmės. HTK sistemoje modelių failai turi labai griežtai apibrėžtą struktūrą. Žemiau pateikiamas pavykslėlis, kaip atrodo „*adomaviciute inga*“ modelių failas *hmm\_adomaviciute\_inga*.

```

~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_E_D_A>
~h "hmm_adomaviciute_inga"
<BEGINHMM>
<NUMSTATES> 19
<STATE> 2
<MEAN> 39
-1.758159e+001 -5.342542e+000 1.763493e+000 -6.846874e+000 -3.253930e+000
-4.135824e+000 -4.317322e+000 -1.382574e+000 -3.843851e+000 1.870433e+000
-3.095777e+000 -1.292532e+000 5.892968e-001 1.147728e-001 3.312866e-002
1.481184e-001 -4.567951e-002 -1.023820e-001 -1.813995e-002 -6.388362e-002
-3.263520e-002 -9.886722e-002 1.778043e-002 -1.287158e-002 -9.484595e-005
4.491300e-003 5.834969e-003 -2.595324e-003 -5.622777e-003 -8.585635e-004
-3.505924e-003 3.840423e-003 -4.847412e-003 -1.435743e-003 5.031886e-003
7.645930e-004 -4.996357e-003 -2.242588e-003 -8.363164e-005
<VARIANCE> 39
6.393353e+001 2.670098e+001 5.863576e+001 4.715696e+001 4.164955e+001
2.867308e+001 3.633283e+001 1.968971e+001 3.282423e+001 1.841615e+001
1.615247e+001 1.117070e+001 6.169783e-002 3.516798e+000 2.436311e+000
3.342489e+000 3.674734e+000 2.795227e+000 2.461520e+000 3.451893e+000
1.747014e+000 2.673193e+000 1.746904e+000 1.618569e+000 1.226119e+000
5.755700e-003 4.768905e-001 4.300175e-001 5.857072e-001 6.417455e-001
4.962888e-001 4.679656e-001 6.011271e-001 3.391850e-001 4.458117e-001
3.594452e-001 3.085575e-001 2.522545e-001 9.844126e-004
<GCONST> 9.804691e+001
<STATE> 3
<MEAN> 39
-1.259330e+001 -9.024642e-001 6.303460e+000 -8.919173e+000 -1.006859e+001
-2.772853e+000 -3.565901e+000 -3.179542e+000 -6.649266e+000 4.419340e+000
-5.703646e+000 -9.832929e-001 6.257125e-001 2.309471e-001 -1.751765e-002

```

4 pav. Pavardės ir vardo „adomavičiūtė inga“ modelių failas

Bendras modelio aprašas pateikiamas šiose eilutėse:

~o – pradžios simbolis;

<STREAMINFO> 1 39 – skaičius 39 nurodo, kad bus naudojamas 39 požymių rinkinys, o skaičius 1 – kad jie bus išdėstyti vienoje eilutėje;

<VECSIZE> 39<NULLD><MFCC\_E\_D\_A> – šioje eilutėje nurodomas vektoriaus dydis, t.y. tie patys 39 požymiai, MFCC\_E\_D\_A reiškia, kad bus naudojami E – energijos, D – delta ir A – delta delta požymiai. Tai dažniausiai naudojamas požymių rinkinys.

~h „hmm\_adomaviciute\_inga” – tai modelio pavadinimas;

<BEGINHMM> – pradžios žyma, atitinkamai pabaigoje bus <ENDHMM> žyma, kuri reikš modelio pabaigą;

<NUMSTATES> 19 – nurodo, kad modelis bus sudarytas iš 19 būsenų. Rekomenduojama būsenų skaičių parinkti lygų garsų skaičiui žodyje. Šiuo atveju pavardė ir vardas „adomaviciute\_inga” turi 17 garsų, o modeliuojamame atvejyje nurodoma 19 garsų. Tai reiškia, jog šio atvejo metu bus pridama prie 17 garsų 2 būsenos.

Toliau aprašomas paties modelio modeliavimas:

<STATE> 2 – pradedama nuo antros būsenos.

<MEAN> 39 – požymių vidurkiai, parenkami intuityviai, modelių mokymo metu bus perskaičiuojami automatiškai, pvz.:

-1.758159e+001 -5.342542e+000 1.763493e+000 -6.846874e+000 -3.253930e+000 -4.135824e+000  
-4.317322e+000 -1.382574e+000 -3.843851e+000 1.870433e+000 -3.095777e+000 -1.292532e+000  
5.892968e-001 1.147728e-001 3.312866e-002 1.481184e-001 -4.567951e-002 -1.023820e-001 -  
1.813995e-002 -6.388362e-002 -3.263520e-002 -9.886722e-002 1.778043e-002 -1.287158e-002 -  
9.484595e-005 4.491300e-003 5.834969e-003 -2.595324e-003 -5.622777e-003 -8.585635e-004 -  
3.505924e-003 3.840423e-003 -4.847412e-003 -1.435743e-003 5.031886e-003 7.645930e-004 -  
4.996357e-003 -2.242588e-003 -8.363164e-005

<VARIANCE> 39 – požymių dispersijos, taip pat parenkamos intuityviai, apmokymo metu bus perskaičiuotos ir bus ištaisytos klaidos, pvz.:

6.393353e+001 2.670098e+001 5.863576e+001 4.715696e+001 4.164955e+001 2.867308e+001  
3.633283e+001 1.968971e+001 3.282423e+001 1.841615e+001 1.615247e+001 1.117070e+001  
6.169783e-002 3.516798e+000 2.436311e+000 3.342489e+000 3.674734e+000 2.795227e+000  
2.461520e+000 3.451893e+000 1.747014e+000 2.673193e+000 1.746904e+000 1.618569e+000  
1.226119e+000 5.755700e-003 4.768905e-001 4.300175e-001 5.857072e-001 6.417455e-001  
4.962888e-001 4.679656e-001 6.011271e-001 3.391850e-001 4.458117e-001 3.594452e-001  
3.085575e-001 2.522545e-001 9.844126e-004

<GCONST> 9.804691e+001 – būsenos stiprinimo koeficientas, parenkamas nuo 0 iki 1.

Analogiški veiksmai atliekami ir su likusiomis būsenomis iki paskutiniosios. Tokiu būdu apskaičiuojamos tikimybės būti pirmoje, antroje, trečioje ir t.t. būsenose. Šiuo atveju, paskutinė yra



19-oji būseną, ji aprašoma kiek kitaip nei ankstesnės, joje sudaroma tikimybių matrica, kuri atrodo taip:

```
0.000000e+000 1.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000
0.000000e+000 9.755885e-001 2.441152e-002 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000
0.000000e+000
```

Eilučių ir stulpelių skaičius turi sutapti. Kadangi priimta, kad tikimybės judės tik į priekį, veiksmas „atgal“ nėra galimas, vadinasi, kiekvienoje eilutėje aprašoma tikimybė likti toje pačioje būsenoje arba pereiti į sekančią (pvz.: pasilikti pirmoje arba pereiti į antrą būseną ir t.t.). Modelio apmokymo metu šios tikimybės taip pat perskaičiuojamos automatiškai.

Pasidarius vieną modelio failą, patogų jį naudoti kaip šabloną kitam modeliui. Reikia tik atitinkamai modifikuoti modelio failą perrašant pavadinimą, būsenų skaičių ir pridėdant arba pašalinant atitinkamas būsenas ir tikimybių matricos eilutes bei stulpelius, priklausomai nuo žodyje ar junginyje esančio garsų skaičiaus (jei tyrime naudojamas žodžių junginys, reikia įskaičiuoti ir brūkšnelius, kurie sujungia žodžius. Pavadinimo pradžia „hmm\_“ neįskaičiuojama).

Atitinkamai reikia modifikuoti tokį patį failą, kai norima, jog 19 garsų turėtų ne tik papildomas būsenas, bet ir Gauso skirstinių modelius, kurių įtaka kalbos atpažinimui yra akivaizdi. Šis metodas priklauso stochastiniam modeliavimui ir modeliuoja statistinį požymių kitimą. Tai statistinis pavaizdavimas, kaip kalbėtojas taria garsus. Gausinis kalbančiojo modelis kalbančiojo požymių pasiskirstymą apibūdina vidurkio vektoriumi ir kovariacine matrica. Norint sukurti failą, kuris turėtų ne tik papildomas būsenas, bet ir modelius, reikia prie kiekvienos būsenos prirašyti Gauso skirstiniams būdingus parametrus. Žemiau pateikiamas paveikslėlis, kaip atrodo balso komandos „*adomaviciute inga*“ modelių failas *hmm\_adomaviciute\_inga*, kai modelį sudaro 4 papildomos būsenos ir 6 Gauso skirstiniai. Atitinkamai po kiekviena būseną prirašomi papildomi parametrai.

```

~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_E_D_A>
~h "hmm_adomaviciute_inga"
<BEGINHHM>
<NUMSTATES> 21
<STATE> 2
<NUMMIXES> 6
<MIXTURE> 1 3.332930e-001
<MEAN> 39
-1.200895e+001 -7.191125e+000 -3.885982e+000 -9.336278e+000 -8.812724e+000
-1.872428e+000 -6.531411e+000 -5.784473e+000 -1.996033e-001 -1.978017e+000
-4.306974e+000 -1.034440e+000 6.917618e-001 2.233078e-002 1.698668e-002
-1.826635e-001 -5.650641e-001 -7.086720e-002 9.852153e-002 1.681378e-001
1.401419e-001 8.040443e-002 -1.133266e-001 6.557339e-002 2.536967e-001
1.213921e-002 -3.569229e-001 -8.108958e-002 -3.470955e-002 1.432953e-001
2.292670e-001 1.795671e-001 3.517151e-001 1.859662e-001 1.258498e-001
1.075179e-001 8.574655e-002 -1.999476e-002 4.217774e-003
<VARIANCE> 39
9.908876e+000 4.669998e+001 4.192262e+001 4.858958e+001 3.755423e+001
4.777534e+001 5.448386e+001 4.320935e+001 3.449192e+001 3.173147e+001
2.851968e+001 2.568670e+001 2.776042e-002 3.396308e+000 3.649181e+000
3.594392e+000 5.802709e+000 4.098303e+000 3.405226e+000 4.605899e+000
3.446178e+000 2.266382e+000 2.111401e+000 2.073932e+000 1.769785e+000
3.594520e-003 3.043140e-001 4.223691e-001 3.633999e-001 5.597968e-001
3.936266e-001 4.317550e-001 6.256956e-001 5.495612e-001 4.172988e-001
4.019326e-001 3.661565e-001 2.902571e-001 3.884512e-004
<GCNST> 1.005290e+002
<MIXTURE> 2 1.092207e-001
<MEAN> 39
-2.404447e+001 3.579013e-002 -1.716035e+000 -8.805790e-001 2.212914e-001

```

**5 pav.** Balso komandos „adomavičiutė inga” modelių failas (+4 būsenos ir 6 skirstiniai)

Penktasis etapas yra modelių apmokymas. Sukurtus modelių failus patogu įsirašyti į atskirą katalogą, pvz. pavadinimu „*models*”. Toliau kuriami *script-failai*, kuriuose surašomi tam žodžiui trečiame etape apskaičiuoti požymių failai. Svarbu nesumaišyti su jau anksčiau aprašytu failu *failai.txt*, kurio turinys yra šiek tiek panašus, tačiau šiame faile reikalinga nurodyti tik požymius ir taką, kaip jie bus pasiekiami. Kiekvienam skirtingam žodžiui kuriamas atskiras tokio pobūdžio failas. Žemiau pateikiami visų tyrimo metu sukurtų pavadinimų pavyzdžiai.

*Trans\_adomaviciute\_inga.scp*

*Trans\_aleksandraviciute\_ramute.scp*

*Trans\_alisauskiene\_jadvyga.scp*

*Trans\_bulachiene\_danguole.scp*

*Trans\_cizaitė\_radvile.scp*

*Trans\_danilkevius\_rytis.scp*

*Trans\_eidziunas\_aurimas.scp*

*Trans\_grigaraviciene\_jurate.scp*

*Trans\_ignatavicius\_gintaras.scp*

Toliau pateikiamas vardo ir pavardės „*adomaviciute inga*“ tokio pobūdžio turinys.

D:\DARBAS\HTK\HTK\pozymiai\0A00.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A01.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A02.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A03.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A04.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A05.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A06.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A07.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A08.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A09.mfc

D:\DARBAS\HTK\HTK\pozymiai\0A10.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A11.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A12.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A13.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A14.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A15.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A16.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A17.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A18.mfc  
D:\DARBAS\HTK\HTK\pozymiai\0A19.mfc

Parengus šiuos failus atliekamas žodžio modelio apmokymas, pasinaudojant *HRest.exe* programa. Kartu su ja komandinėje eilutėje užrašomi tokie parametrai: *HRest.exe -T 1 -S trans\_adomaviciute\_inga.scf -L labels\models\hmm\_adomaviciute\_inga*

Čia *trans\_adomaviciute\_inga.scf* yra script failas, kuriame išvardinti žodžių junginio požymių failai, po parametro 1 rašoma duoto žodžio, kuriam atliekamas apmokymas pavadinimas taip, kaip buvo užrašyta žymių faile, po to nurodomas katalogas, kur yra surašyti duoto žodžio žymių failai, galiausiai rašomas katalogas ir modelio failo pavadinimas. Atlikę tokį veiksmą gauname apmokytą modelio failą, kurį galima naudoti atpažinimo metu. Toks veiksmas atliekamas visiems turimiems žodžiams, komandinėje eilutėje atitinkamai pakeičiant tik žodžio ar žodžių junginio pavadinimą.

Šeštasis etapas yra testavimas. Paskutiniam etapui reikalinga paruošti naujus tų pačių žodžių įrašus. Šiam tyrimui imami kitų diktorių *fiveval* ir *mkasrat* paruošti tokie patys garso įrašai po 20 ištarimų. Paruošiamas dar vienas script-failas, pavadinimu *test.scf*, kurio turinys atrodo taip, kaip 3 etape kurto failo *failai.scf*, pakoreguojant taką, kuriuo bus pasiekiami nauji įrašai. Sukūrus šį failą reikia apskaičiuoti požymius, kaip buvo daroma 3 etape. Komandinėje eilutėje rašomas toks tekstas: *Hcopy.exe -C config -S test.scf*. Apskaičiuoti požymių failai surašomi į dar vieną *script-failą*, kaip buvo daroma 5 etape, tik šį kartą požymiai nurodomi viename faile. Failas pavadintas *test\_mfc.scf*. Ištrauka iš jo turinio atrodo taip:

D:\DARBAS\HTK\HTK\irasai2\8A00.wav D:\DARBAS\HTK\HTK\irasai2\8A00.mfc  
D:\DARBAS\HTK\HTK\irasai2\8A01.wav D:\DARBAS\HTK\HTK\irasai2\8A01.mfc  
D:\DARBAS\HTK\HTK\irasai2\8A02.wav D:\DARBAS\HTK\HTK\irasai2\8A02.mfc  
...  
D:\DARBAS\HTK\HTK\irasai2\8B00.wav D:\DARBAS\HTK\HTK\irasai2\8B00.mfc  
D:\DARBAS\HTK\HTK\irasai2\8B01.wav D:\DARBAS\HTK\HTK\irasai2\8B01.mfc  
D:\DARBAS\HTK\HTK\irasai2\8B02.wav D:\DARBAS\HTK\HTK\irasai2\8B02.mfc  
D:\DARBAS\HTK\HTK\irasai2\8B03.wav D:\DARBAS\HTK\HTK\irasai2\8B03.mfc  
...

Toliau kuriamas modelių sąrašo failas, į kurį surašomi visi modelių failų pavadinimai. Pačiam failui suteikiamas "*hmmlist*" pavadinimas. Jo turinys atrodo taip:

*hmm\_adomaviciute\_inga*

*hmm\_aleksandraviciute\_ramute*

*hmm\_alisauskiene\_jadvyga*

*hmm\_bulachiene\_danguole*

*hmm\_cizaitė\_radvile*

*hmm\_danilkevicius\_rytis*

*hmm\_eidziunas\_aurimas*

*hmm\_grigaraviciene\_jurate*

*hmm\_ignatavicius\_gintaras*

*hmm\_ivanceviciute\_erika*

Priešpaskutinis žingsnis yra žodyno failo kūrimas. Failas pavadinamas *dict.txt*, o jo turinys atrodo taip:

*adomaviciute\_inga*

*aleksandraviciute\_ramute*

*alisauskiene\_jadvyga*

*bulachiene\_danguole*

*cizaitė\_radvile*

*danilkevicius\_rytis*

*eidziunas\_aurimas*

*grigaraviciene\_jurate*

*ignatavicius\_gintaras*

*ivanceviciute\_erika*

*hmm\_adomaviciute\_inga*

*hmm\_aleksandraviciute\_ramute*

*hmm\_alisauskiene\_jadvyga*

*hmm\_bulachiene\_danguole*

*hmm\_cizaitė\_radvile*

*hmm\_danilkevicius\_rytis*

*hmm\_eidziunas\_aurimas*

*hmm\_grigaraviciene\_jurate*

*hmm\_ignatavicius\_gintaras*

*hmm\_ivanceviciute\_erika*

Šio failo paskirtis – pranešti kompiuteriui, ką reikia išvesti į ekraną, jei bus atpažintas vienas ar kitas žodis (modelis), pvz., jei panašiausias modelis bus *hmm\_adomaviciute\_inga*, tai kompiuteris atspausdins tekstą „*adomaviciute inga*“ ir t.t.

Paskutinis etapas yra testavimas. Tam naudojama programa pavadinimu *HVite.exe*, komandinėje eilutėje surenkant tokią komandą: *HVite.exe -T 1 -S test\_mfc.scf -i results -w wordnet.txt dict hmmlist*.

Čia *test\_mfc.scf* yra script failas, kurtas 6 etape, *wordnet.txt* – tai žodžių tinklas, sukurtas 2 etape, *dict* failas – priešpaskutiniame žingsnyje kurtas žodyno failas, *hmmlist* – 6 etape kurtas modelių sąrašo failas, o *results* – failas, kuriame gaunami testavimo rezultatai.

### 3.4.1. Pavardžių ir vardų tyrimų su HTK rezultatų analizė

Įgyvendinus visus anksčiau aprašytus etapus, po testavimų sugeneruojamas failas *results*, kuriame yra pateikiami atpažinimo rezultatai, todėl siekiant išsiaiškinti atpažinimo kokybės tikslumo procentą privaloma išanalizuoti gautus rezultatus. Rezultatai faile yra išspausdinami tvarka, kuria įrašai pateikiami faile *test\_mfc.scf*, tai reiškia, jog tyrimo atveju, pirmieji 20 rezultatų gauti balso komandai „*adomaviciute inga*“, antrieji 20 – „*aleksandraviciute ramute*“, tretieji – „*alisauskiene jadvyga*“ ir t.t.

Norint pasiekti geriausių vardų ir pavardžių atpažinimo rezultatus, kai tyrimai yra atliekami su HTK, reikia tinkamai parinkti parametrus, galinčius padėti pasiekti užsibrėžtą tikslaus kalbos atpažinimo tikslą. Kaip jau buvo sakyta, kiekviena balso komanda buvo modeliuojama viena ištinine paslėpta Markovo grandine. Būsenų skaičius grandinėje buvo parinktas apytikriai lygus balso komandą sudarančių fonetinių elementų skaičiui ir svyravo tarp 15 ir 24. Mažiausiai – 15 būsenų –

buvo naudotos modeliuoti komandą „*cizaitė radvilė*“, daugiausiai – 24 būsenomis – modeliuota balso komanda „*aleksandraviciute ramute*“. Visi eksperimentai atlikti tokiu principu: 8 diktorių, t.y. *fdeilei, fginbar, mdomvas, malgzil, frimram, fievzil, fraival, mvilram* (moteriški ir vyriški balsai) ištarimai (po 20 kiekvienos komandos ištarimų) buvo naudojami apmokymui, o 2 diktorių, t.y. diktorės *fiveval* bei diktoriaus *mkasrat* balsais įdiktuoti įrašai naudoti testavimui.

Taigi vardų atpažinimo rezultatai pirmiausia buvo gauti tada, kai tyrimui būsenų skaičius grandinėje buvo parinktas lygus balso komandą sudarančių fonetinių elementų skaičiui (0 būsenų). Atpažinimo tikslumas tenkinant minėtas sąlygas yra tik **61,5 proc.** Siekiant tikslesnio atpažinimo, prie vadinamųjų 0 būsenų (kai būsenų skaičius atitinka fonetinius žodžio elementus) buvo pridėtos 2 būsenos. Tai reiškia, jog jeigu prie mažiausiai raidžių turinčios komandos „*cizaitė radvilė*“ (14 raidžių + tarpas tarp vardo ir pavardės = 0 būsenų) bus pridėtos +2 būsenos, tai iš viso „*cizaitė radvilė*“ turės 17 būsenų. Atitinkamai „*adomaviciute inga*“ (17 raidžių = 0 būsenų) turės 19 būsenų ir pan. Pridėjus kiekvienam vardui po 2 būsenas, rezultatas, atlikus tyrimą gaunamas šiek tiek geresnis – **67,5 proc.** Vėliau tyrimas kartojamas su +4, +5 ir +6 būsenomis, nes manoma, jog didinant būsenų skaičių gerės balso komandų atpažinimas. Geresnis rezultatas buvo gautas pridėjus 4 būsenas, šios sąlygos atpažinimo tikslumas siekia **80,5 proc.**, esant sąlygai +5 būsenos, tikslumas sumažėjo iki **67,25 proc.** O esant pastarajai sąlygai +6 būsenos, atpažinimo tikslumas vėl padidėjo – **77,5 proc.** Būsenų didinimas šiuo atveju buvo naudingas tik iš dalies, nes tyrimui naudojami vardai yra ilgi, todėl didelis būsenų skaičius ilgiems vardams yra per didelis. Dėl šios priežasties sugeneruojama daug vidinių parametrų failuose, tai turi įtakos kalbos atpažinimo tikslumui, tačiau tyrimų kiekis leido pastebėti šią tendenciją. Tyrimo priklausomybę nuo būsenų skaičiaus galima pamatyti 2 priede.

Siekiant gauti geresnius rezultatus testuojant tą patį pavardžių ir vardų garsyną, reikia keisti parametrus. Kadangi atliekant tyrimus su būsenomis geriausias rezultatas buvo gautas esant sąlygai +4 būsenos, tolimesniems tyrimams parenkami parametrai turės +4 būsenas ir atitinkamą 2 ir 6 Gauso skirstinių skaičių. Tikėtina, kad tai gali pagerinti kalbos atpažinimo rezultatus. Gauso skirstinių pridėjimas prie būsenų skaičiaus nepatvirtino hipotezės apie geresnį kalbos atpažinimo rezultatą. Pirmuoju atveju (+4 būsenos ir 2 skirstiniai) gaunamas **65,5 proc.** atpažinimo tikslumas, nes ne tik dvi iš balso komandų atpažįstamos idealiai, likusios siekia menką atpažinimo tikslumą. Šiuo atveju žemas atpažinimo tikslumas gaunamas todėl, kad būsenų pasiskirstymas pridėjus kiekvienai balso komandai po 4 būsenas svyruoja nuo 19 būsenų iki 28 būsenų. Tikėtina, kad dėl skirtingų būsenų skaičiaus komandose ir žemo Gauso skirstinių skaičiaus ilgomis balso komandoms padarė įtaką kalbos atpažinimui ir nebuvo pagerintas komandų atpažinimas. Antruoju atveju – +4 būsenos ir 6 skirstiniai prie tos pačios imties būsenų, kaip ir pirmuoju atveju, pridėdama daugiau parametrų (Gauso skirstinių), kas, tikėtina, šiuo atveju padės išgauti didesnę atpažinimo rezultatą dėl

didelio parametrų skaičiaus. Antruoju atveju, kaip ir pirmuoju, vis dėlto buvo gautas žemas atpažinimo rezultatas – **63,25 proc.**, žemesnis nei esant 2 Gauso skirstiniams. Analizuojant rezultatus detaliau pastebima, kad mažesnę būsenų skaičių turinčios balso komandos abiem atvejais, kai buvo pridėti skirstiniai, atpažintos geriau nei tuo atveju, kai būsenų skaičius buvo didesnis (žr. 3 priede). Dėl šios priežasties buvo nuspręsta atlikti keletą tyrimų su fiksuotu būsenų ir Gauso skirstinių skaičiumi visoms balso komandoms, tokiu būdu siekiant geresnio kalbos atpažinimo rezultato. Pirmiausia visoms balso komandoms buvo nustatytas fiksuotas 14 būsenų ir 4 mišinių skaičius. Tai reiškia, kad balso komanda, įprastai turinti, pavyzdžiui, 15 būsenų („*cizaitė radvilė*“) turės 14 būsenų, kas reiškia, jog būseną bus atpažįstama kaip „*cizaitė radvilė*“ ir pan. Atlikus tyrimą atpažinimo gautas rezultatas – **61,75 proc.** (žr. 4 priedas). Kitu atveju, esant fiksuotoms 14 būsenų ir 16 skirstinių skaičiui pasiektas labai aukštas atpažinimo rezultatas, kuris siekia **96,75 proc.** Antruoju atveju rezultatas gautas labai aukštas, palyginus su pirmuoju, nes fiksuotas būsenų skaičius yra gana mažas, o skirstinių kiekis yra per didelis, lyginant šiuos parametrus su balso komandų ilgiais. Paliekant tokį patį Gauso skirstinių skaičių, t.y. 4 ir 16 bus atlikta dar keletas tyrimų, siekiant nustatyti geriausią parametrų variantą, kurį nustačius pasiekiamas geriausias kalbos atpažinimo rezultatas. Nustačius fiksuotą būsenų skaičių 16, prie jų dar pridėjus 4 Gauso skirstinius kalbos atpažinimo rezultatas gautas **70 proc.** (žr. 5 priedas). Antruoju bandymu, kai fiksuotas būsenų skaičius yra 16 ir toks pats Gauso skirstinių skaičius, atpažinimas ženkliai pagerėjo – **96,75 proc.** Tyrimai su skirtingais parametrais parodė, kad esant fiksuotam būsenų skaičiui ir pridedamiems Gauso skirstiniams, gaunamas geresnis rezultatas nei pirmaisiais atvejais, kai būsenų skaičius tarp balso komandų buvo skirtingas. Taip yra dėl to, kad HTK programų paketui lengviau išanalizuoti duomenis su jiems pritaikytais parametrais, dėl ko ir atpažinimo tikslumas gaunamas geresnis, nei esant skirtingiems parametrams.

Atlikus eilę tyrimų su fiksuotu būsenų skaičiumi ir Gauso skirstiniais bei gavus geriausius rezultatus esant fiksuotoms 14 – ai būsenų ir 16 skirstinių, bei 16 – ai būsenų ir 16 skirstinių – po **96,75 proc.** tikslumą, nuspręsta palikti 16 skirstinių (su jais gautas geriausias rezultatas) ir didinti fiksuotą būsenų skaičių, siekiant nustatyti ribą, kada bus gautas aukščiausias atpažinimo rezultatas. Nustačius reikiamą ribą vėliau bus galima atlikti tolimesnius tyrimus panaudojant netaisyklingai ištartas balso komandas. Taigi paliekant Gauso skirstinių skaičių 16 parenkamas fiksuotas būsenų skaičius 17 ir 21. Pirmu atveju sugeneruotas rezultatas yra **92,5 proc.**, antruoju – **81 proc.** Didelis būsenų skaičius ir didinamas fiksuotų būsenų skaičius nedavė dar aukštesnių rezultatų nei prieš tai. Vadinasi, tolimesniems tyrimams bus panaudoti parametrai, kurie buvo sugeneravę iki šiol aukščiausią rezultatą – **96,75 proc.** (žr. 6 priedas).

Taip pat surastos vidutinės gerai ir blogai ištartų komandų atpažinimo su HTK paketu logaritminės tikimybės. Pirmuoju atveju logaritminė tikimybė yra ~ -12047 antruoju ~ -17157. Pagal

jas nustatytas slenkstis yra  $\sim -14602$ . Logaritminės tikimybės, gautos su HTK paketu, turi žymiai geresnę skiriamąją galią nei atpažinimo matas, gautas su ispanų kalbos atpažintuvu. Tokiu atveju galimas sprendimas būtų realizuoti hibridinį atpažintuvą, apjungiant abu atpažintuvus ir pagerinant atpažinimo tikslumą, o slenkstį realizuoti tik pagal HTK paketo pateikiamas logaritmines tikimybes. Tai leistų pasiekti aukštą balso komandų atpažinimo tikslumą [22].

Galiausiai kaip ir su ispanų kalbos atpažintuvo apmokytu profiliu *Iveta1*, taip ir su HTK atliekamas tyrimas, kurio metu testuojamas netaisyklingai ištartų balso komandų garsynas. Šiam tyrimui bus panaudotos dvi skirtingos sąlygos, t.y. fiksuotas būsenų skaičius 14 ir 16 mišinių, ir fiksuotas būsenų skaičius 16 ir 16 mišinių. Kaip ir pastaraisiais atvejais, pritaikius minėtuosius parametrus buvo gauti aukščiausi balso komandų atpažinimo rezultatai. Pirmuoju atveju atpažinimo rezultatas yra **47 proc.**, antruoju – **44 proc.** Nepaisant nustatytų parametrų, kurie turėtų sugeneruoti aukštą atpažinimo rezultatą, rezultatas buvo gautas prastesnis nei tas, kurį sugeneravo ispanų kalbos atpažintuvus. Vadinasi, atpažįstant iškraipytas balso komandas, geresnis rezultatas gaunamas su ispanų kalbos atpažintuvu, todėl tikėtina, kad programa, besiremianti ispanų kalbos atpažintuvu, blogai tariamas komandas atpažins tiksliau.

### 3.5. Tyrimų dalies išvados

1. Baigiamajam magistro darbui realizuoti naudojami metodai: deskriptyvinis, eksperimentinis ir kiekybinės analizės.
2. Magistro darbo metu sukurti specialūs daugiadiktoriniai garsynai: taisyklingai ištartų sudėtingų lietuviškų pavardžių ir vardų bei netaisyklingai tariamų tų pačių pavardžių ir vardų. Atlikus garsynų tyrimus nustatyta balso komandų atpažinimo kokybė: su apmokytu ispanų kalbos profiliu *Iveta1* taisyklingai ištartas balso komandas atpažintuvus atpažįsta **98,4 proc.** tikslumu, o netaisyklingai ištartas – **58,4 proc.**
3. Atlikti tyrimai su HTK programų paketu parodė, kad taisyklingai ištartas balso komandas atpažintuvus atpažįsta esant parametrams: fiksuotas būsenų skaičius 14 ir 16 skirstinių – **96,75 proc.**, o netaisyklingai ištartas – **47 proc.** Tikslumu, esant fiksuotam būsenų skaičiui 14 ir 16 skirstinių.
4. Gautais tyrimų rezultatais tikslinga remtis kuriant balso atpažinimo programą ir atrenkant balso komandas, kurias ji atpažins.
5. Nustatyti ispanų kalbos atpažintuvu ir HTK programų paketu atliktų tyrimų slenksčiai. Ispanų kalbos atpažintuvu atliktų tyrimų slenkstis yra 533,754. HTK paketo programomis atliktų tyrimų slenkstis yra  $\sim -14602$ .

## 4. PROJEKTINĖ DALIS

Užsienio studentų lietuvių kalbos mokymosi problema turi būti sprendžiama ugdant visas kalbines kompetencijas. Siekiant tokio tikslo, kalbiniams įgūdžiams formuoti bus pasitelkiamas žmogaus kalbos atpažinimas, kuris bus pritaikytas prie mokymosi medžiagos. Medžiaga bus pateikta mokymosi kurse VMA naudojant įdomius metodus ir panaudojant inovatyvius sprendimus.

Kalbos atpažinimo technologijų panaudojimas mokantis kalbos privalo neišsiskirti iš bendro konteksto. Tai reiškia, kad naudojantis įvairiais šaltiniais bus sukurta lietuvių kalbos mokymosi sistema, pritaikyti metodai medžiagai įsisavinti ir kalbiniams įgūdžiams formuoti.

Šiame skyriuje aprašomos dėstytojų ir studentų apklausos, kuriomis remiantis kuriama kurse naudojama mokymo metodika bei pateikiama medžiaga. Taip pat pateikiama kuriamo kurso sistema ir sistemą sudarantys komponentai.

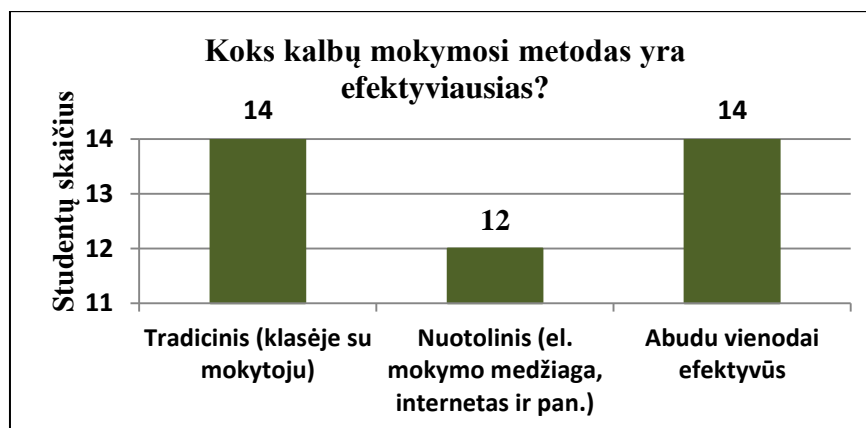
### 4.1. Lietuvos studentų poreikių analizė

Kiekvienos kalbos mokymasis turi bendrus principus, remiasi panašiais metodais. Dėl šios priežasties siekiant ištirti studentų poreikį studijuoti svetimą užsienio kalbą virtualiu būdu bei išsiaiškinti priimtinausius kalbos mokymosi būdus pirminė analizė atlikta su Lietuvos studentais, su jais yra lengvesnė susiekimo bei bendradarbiavimo galimybė. Apklausa buvo parengta apklausa.lt tinklalapyje, kuria buvo apklausti 40 studentų iš skirtingų universitetų. Apklausoje anketą sudaro 14 klausimų, kurie yra bendro pobūdžio (lytis, amžius, mokymosi įstaiga ir pan.) ir susiję su informacijos rinkimu apie užsienio kalbos mokymą virtualiu būdu. Apklausa buvo platinama socialiniame tinkle „Facebook“.

Šiame poskyryje bus glaustai aptarti svarbiausi apklausoje klausimai. Detalią apklausoje analizę galima rasti 15 priede. Apklausoje savo nuomonę pareiškė 18 moterų (45 %) ir 22 vyrai (55 %). Dauguma iš apklaustųjų (19 stud., 47,5 %) nurodė esantys iš KTU, 10 (25 %) nurodė savo mokslo įstaigą esant VDU. Likusieji pasiskirstė taip: po 5 asmenis (12,5%) iš VGTU ir VU, ir 1 (2,5%) nurodė, jog yra LSU studentas.

Svarbiu laikytas klausimas, kuriuo siekiama sužinoti, koks kalbų mokymosi metodas studentams atrodo efektyviausias. Apklausoje dalyvavusiems studentams labiausiai priimtinais kalbos mokymosi metodais išlieka visiems gerai žinomas tradicinis metodas, kai mokytojas perduoda studentams informaciją per tiesioginius susitikimus, moko ją pritaikyti (14 stud.). Mažiau patikimu laikytinas nuotolinio mokymosi metodas, kai mokymasis paremtas savarankišku darbu, naudojant elektroninę mokymosi medžiagą, skaidres, internetą ir kt. Likusieji 14 studentų (kaip ir pirmuoju atveju) pasisako už tai, jog tiek tradicinis, tiek nuotolinis mokymas leidžia pasiekti vienodus rezultatus, nes abu yra laikomi lygiaverčiais. Pagal pateiktą diagramą matoma, jog studentų nuomonė šiuo klausimu buvo skirtinga.

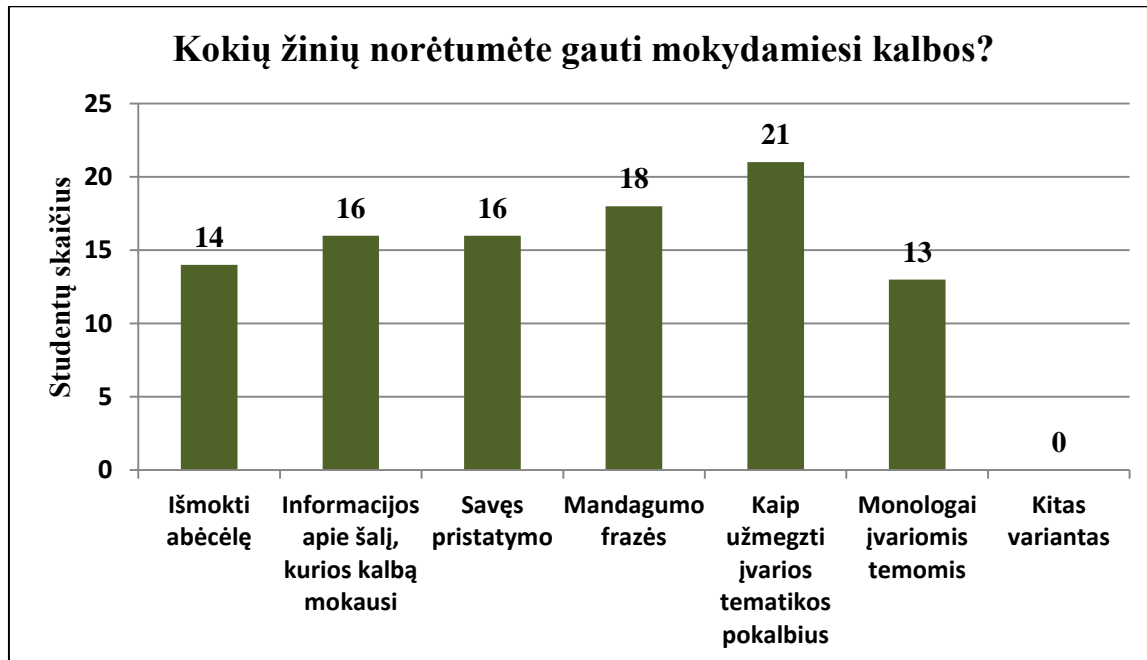




6 pav. Efektyviausi kalbų mokymosi metodai

Į klausimą, kokio tipo kalbos įgūdžių formavimas studentams yra svarbiausias (kokius įgūdžius norėtų suformuoti), dauguma atsakė, jog svarbiausiais išlieka kalbėjimo ir rašymo įgūdžiai, o mažiausiai svarbiu įvardijo gramatikos suvokimą.

Kitu klausimu dauguma (21 studentas) pasisakė, jog aktualiausia būtų išmokti užmegzti pokalbį su kitu žmogumi (buitinei kalbai būdingas bruožas, tad būtų logiška, kad studentas pradeda mąstyti nuo paprastų dalykų išmokimo). Ne mažiau svarbu mokėti elementaraus etiketo frazių, gebėti pristatyti save svetimiems žmonėms ir išmanyti svetimos kultūros peripetijas (informacija apie šalį). Mažiau svarbu studentams pasirodė mokytis abėcėlės (nebent tai būtų rusų kalbos abėcėlė) ar pasisakyti įvairiomis temomis / klausimais (monologai).



7 pav. Žinios mokymosi procese

Pasirinktus dalykus studentai geriausiai išmoktų skaitydami tekstus (27,1 % , 23 stud.), versdami tekstus (24,7% , 21 stud.). Nežymiai, bet mažiau naudos teiktų kalbėjimasis su kitu žmogumi (22,4 % , 19 stud.) ir dialogų rašymas (20 % , 17 stud.).

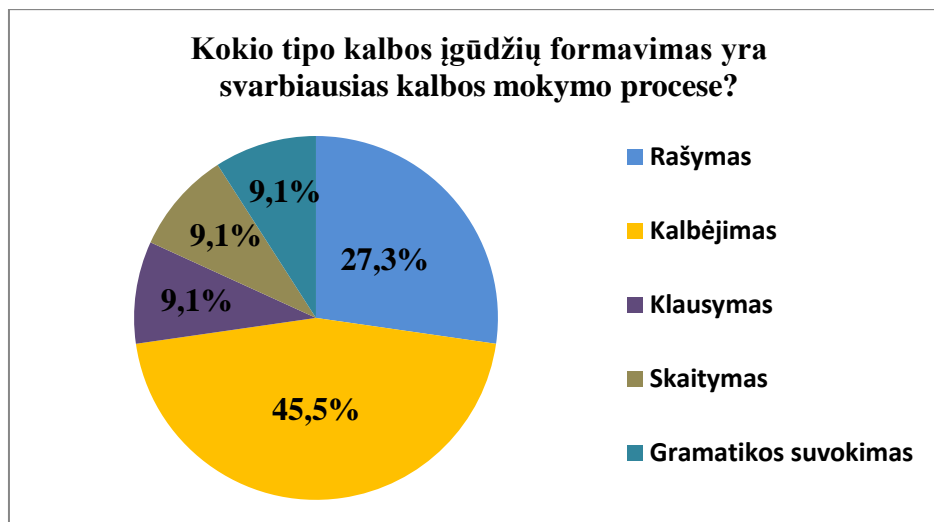
Kalbos mokymo metodika, kai mokymo procese taikomas įvairių užduočių, pateikiamų su atsakymų variantais atlikimas, studentams pasirodė nereikšmingas, todėl sulaukė tik 5 apklaustųjų dėmesio, kas sudaro 5,9%. Paskutiniu metu su kalbos mokymu susijusiu klausimu norėta sužinoti, kokia virtualaus kalbos mokymo metodika studentams atrodo efektyviausia. Prioritetas teikiamas skaidrėms ir vaizdo įrašams. Visa tai galima patalpinti studentų nurodytoje bet kokioje mokymui paskirtoje internetinėje svetainėje (44,4% - 20 stud.) arba VMA „Moodle“ (18 studentų (40%)).

#### **4.2. Dėstytojų apklausos analizė**

Nepaisant to, jog studentai iš patirties gali įvardinti, kokios priemonės ir mokymosi metodai jiems gali būti naudingi mokymosi procese, būtina atkreipti dėmesį ir į ekspertų nuomonę, ką jie gali pasakyti ta pačia tema. Dėl šios priežasties buvo sugeneruota apklausa apklausa.lt tinklalapyje, kuria buvo apklausti 5 KTU lietuvių kalbos specialistai-ekspertai, siekiant išsiaiškinti, kokius, pasak kalbos specialistų (kurie nuolatos susiduria su kalbos mokymu, jame taikomomis priemonėmis), metodus yra tikslingiausia taikyti mokant užsienio studentus lietuvių kalbos virtualiu būdu. Norint nustatyti priimtina ekspertų skaičių, reikia vadovautis metodologinėmis prielaidomis, kurios yra suformuluotos klasikinėje testų teorijoje. Teorijoje teigiama, jog agreguotų sprendimų patikimumą ir priimančių sprendimą (šiuo atveju ekspertų) skaičių sieja greitai gėstantis netiesinis ryšys. Įrodyta, jog agreguotų ekspertinių vertinimų moduliuose su vienodais svoriais nedidelės ekspertų grupės sprendimų ir vertinimų tikslumas nenusileidžia didelės ekspertų grupės sprendimų ir vertinimų tikslumui [2].

Apklausos anketą sudaro 8 klausimai, kurie ne detaliam apsprendžia svarbiausius punktus, į kuriuos reikia orientuotis prieš pradėdant kurti virtualų kalbos mokymosi kursą. Apklausos platinimas vyko naudojantis bakalauro mokymosi metais sukauptais dėstytojų kontaktais.

Tiek studentų, tiek dėstytojų apklausos tikslas yra vienodas, t.y. sužinoti, kokie kalbos mokymosi principai būtų efektyviausi teikiant virtualų kursą. Visų pirma, dėstytojams skirtoje apklausoje klausama, kokio tipo kalbos įgūdžių formavimas yra svarbiausias kalbos mokymo procese. Svarbiausiais įgūdžiais yra laikomi kalbėjimo bei rašymo įgūdžiai. Mažiau svarbiais yra laikomi likusieji įgūdžiai, t.y. klausymo, skaitymo bei gramatikos suvokimo. Skirtingą nuomonių pasiskirstymą įrodanti skritulinė diagrama pateikiama žemiau.



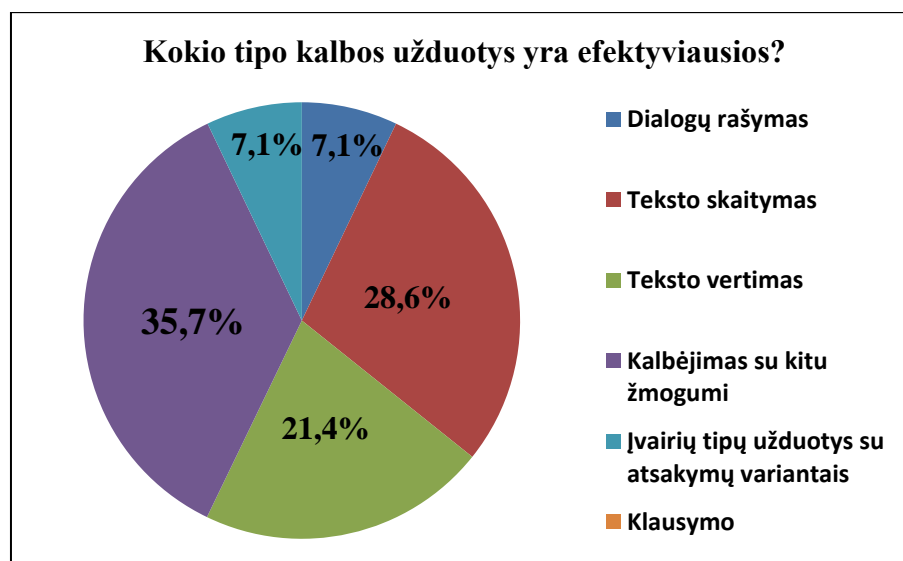
**8 pav.** Įgūdžiai kalbos mokymosi procese

Priemonės, kalbos mokymosi procese teikiančios patrauklumo – vaizdinės medžiagos pateikimas (36,4%). Vienodai patraukliu teikimo būdu (27,3%) yra laikoma teikiamų užduočių įvairovė bei teikimo paprastumas.

Efektyviausia kalbos mokymosi metodika, pasak dėstytojų, remiasi paskaitų vaizdo įrašų ir skaidrių teikimu. Mažiausiai naudos mokymosi procese gali teikti tik skaidrių teikimu paremta mokymo metodika.

Taip pat, kad būtų galima orientuotis, kuria kryptimi vykdyti mokymus, kokios informacijos teikimas kurse būtų tikslingas, buvo klausama, kokios temos yra neišvengiamos kalbos mokyme. Svarbiausiu mokymo aspektu jie išskiria savęs pristatymo, o likusias temas apie abėcėlę, šeimą ir mandagumą paliko antroje vietoje.

Pasirinktus dalykus, pasak dėstytojų, geriausiai studentai įsisavintų, jeigu kalbėtusi su kitais žmonėmis svetima kalba (~36% - 5 dėst.), skaitytų tekstus (~29% - 4 dėst.) bei verstų tekstus (~21% - 3 dėst.).



**9 pav.** Efektyvių kalbos užduočių tipai

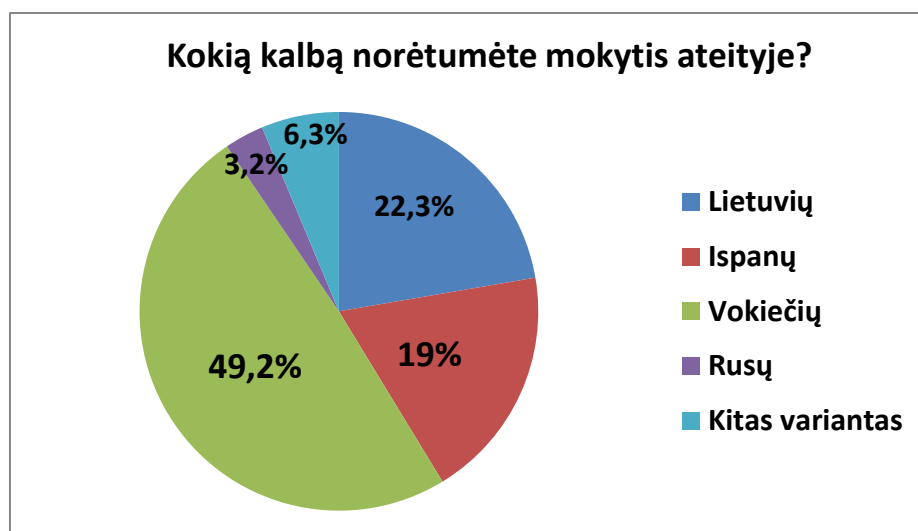
Galiausiai dėstytojai nurodė, jog patogiausia pateikti mokymosi medžiagą studentams yra Moodle aplinkoje.

### 4.3. Užsienio studentų poreikių analizė

Siekiant iširti užsienio studentų poreikį studijuoti svetimą užsienio kalbą virtualiu būdu, buvo specialiai užsienio studentams sugeneruota apklausa anglų kalba [apklausa.lt](http://apklausa.lt) tinklalapyje. Apklausa buvo pateikta praėjusių metų rudens semestre studentams, studijuojantiems Kauno technologijos universitete. Apklausa išplatinta bendradarbiaujant su viena iš KTU universiteto dėstytojų, kuri dėsto paskaitas užsienio studentams. Vienos iš dėstytojos paskaitų metu studentams buvo pateikta užduotis – atlikti apklausą. Visi atsakiusieji į apklausos klausimus yra pirmo kurso studentai, 5 yra Erasmus programoje dalyvaujantys asmenys, o likusieji – nuolatinių KTU universiteto studijų studentai.

Šiame poskyryje bus glaustai aptarti svarbiausi apklausos klausimai. Detalią apklausos analizę galima rasti 19 priede. Atsakymus į apklausos klausimus pateikė 35 užsienio studentai. Šiuo atveju tai yra 29 (82,9%) vyrai ir tik 6 (17,1%) moterys. Didžioji dalis, net 25 (71,4%) studentai, teigia esantys iš Indijos, 3 (8,6%) iš Ispanijos, o likusieji 7 (20%) savo gimtąja šalimi nurodė esant Italiją (2 stud.), Braziliją (1 stud.), Latviją (1 stud.), Azerbaidžaną (1 stud.) ir Libaną (2 stud.).

Užsienio studentų buvo paklausta, kokią kalbą jie norėtų išmokti artimiausioje ateityje. Dauguma studentų, net 31 stud. (49,2%), nurodė pageidaujantys išmokti vokiečių kalbos ypatumų, lietuvių kalbos žinių įgijimas domina 14 (22,3%) apklaustųjų, 12 stud. (19%) domina ispanų kalba, 2 stud. (3,2%) – rusų kalba, o kiti 4 stud. (6,35%) įvardijo tokias kalbas kaip prancūzų, arabų, italų ir portugalų. Šiame klausime buvo galima pažymėti kelis pageidaujamus atsakymo variantus, kas suformavo tokius atsakymo rezultatus:



10 pav. Pageidautina išmokti kalba

Vienodai efektyviais kalbos mokymosi būdais studentai įvardijo nuotolinį ir tradicinį, tai sudarė 51,4%.

Patrauklumo kalbų mokymo medžiagai, pasak studentų, gali suteikti pirmiausia iliustracinė medžiaga (20 stud. – 20,6%), taip pat jos pateikimo ir naudojimo paprastumas (19 stud. – 19,6) bei užduočių įvairovė (19 stud. – 19,6%).

Įgūdžiai, kuriuos norėtų suformuoti besimokantys kalbos studentai – užmegzti pokalbį (20 stud. – 24,1%) bei prisistatyti (19 stud. – 22,9%). Užduotys, kurioms apklaustieji skiria pirmumą yra kalbėjimo užduotys (29 stud. – 34,9%).

Efektyviausias metodas/technika norint išmokti kalbos yra dialogai tarpusavyje (23 stud. – 24,5%), edukaciniai žaidimai (19 stud. – 20,2%) ir vaizdo įrašų/filmų žiūrėjimas (18 stud. – 19,1%).

Kaip ir dėstytojų apklausos atveju, užsienio studentų taip pat buvo klausiama apie patogiausią terpę kalbos medžiagai talpinti. Studentai atsakė lygiai taip pat kaip ir dėstytojai – patogiausia terpė yra „Moodle“, su tuo sutinka didžioji dalis studentų (30 stud. – 45,5%).

Atlikta apklausa parodė, jog į KTU atvykę studentai užsieniečiai yra suinteresuoti išmokti daug naujų kalbų, taip pat ir lietuvių kalbos. Priimtinausi kalbų mokymosi metodai studentams yra tradicinis ir nuotolinis, kurie yra vienodai efektyvūs. Kalbos mokymesi svarbiausia suformuoti kalbinius įgūdžius, kurį galima perteikti medžiagos paprastumu, užduočių įvairove ir vaizdine medžiaga.

#### **4.4. Lietuvių kalbos mokymo kurso metodika**

Kurse kalbai mokyti persipina keletas skirtingų metodų, kurie atsakingi už išmoktus skirtingus kalbinius aspektus. Pagrindiniai kalbos mokymo principai kurse: kalbos mokymas pasitelkiant teorinę kalbos mokymosi medžiagą – vaizdinę, garsinę, – įdomias užduotis ir testus. Visame kurse medžiaga pateikiama dviem kalbomis: lietuvių ir anglų.

Prieš pradėdant mokytis kalbos, ne visi besimokantieji pažįsta šalį, kurios kalbos ketina mokytis. Dėl šios priežasties, pirmiausia kurse naudinga pateikti minimalią informaciją apie šalį: didžiausius miestus, kultūrinius aspektus, papročius ir kt. Tai galima padaryti pateikiant besimokantiems peržiūrėti linksmo turinio vaizdo įrašus. Vaizdo įrašų teikimas derinamas su rašytine medžiaga – pagrindiniais faktais apie šalį bei žodynu, pradinėms žinioms formuoti. Pirminėms žinioms patikrinti taip pat naudojamas įdomus būdas – kryžiažodis. Vėliau šį žinių patikrinimo būdą galima įkomponuoti prie sunkesnio pobūdžio temų.

Lietuvių kalba turi unikalią abėcėlę, todėl pirma tema po šalies pristatymo yra susijusi su abėcėlės mokymu. Kadangi tai tema, nuo kurios priklauso tolimesni kalbos mokymosi rezultatai, svarbu atitinkamai pateikti mokymosi medžiagą ir pateikti jos kuo daugiau ir aiškiau. Tuo remiantis, abėcėlės mokoma naudojant vaizdo įrašus, interaktyvią medžiagą, kurią galima ir perskaityti, ir perklausyti. Taip pat pateikiama teorinė mokymosi medžiaga, paaiškinanti kiekvienos abėcėlės raidės tarimo ypatybes. Privalu kruopščiai atlikti užduotis ir įsigilinti į teorinę medžiagą, nes žinioms patikrinti pateikiamas testas, kuriame klausimai yra susiję su abėcėlės raidžių tarimo ypatybėmis.

Elementarus abėcėlės išmokimas – tai perėjimas sunkesnio mokymosi lygio link. Norint pereiti prie kitos mokymosi temos, tikslinga atlikti visas užduotis iš pradėtos temos.

Leksikos mokymas kurse yra vykdomas prie atitinkamų temų pateikiant žodyną, nepaisant to, į ką tema yra orientuota. Tai daroma pateikiant temoje vyraujančius žodžius temos pabaigoje. Be viso to, leksikos mokymui yra paskirta atskira tema. Leksikos mokymo metodas – skaitymas ir klausymas. Kiekvienoje temoje pateikiamas interaktyvus tekstas su žodžiais, kuriuos galima perklaudyti. Taip lengviau įsiminti pateiktą medžiagą. Antrasis metodas leksikai mokyti – žodžių vertimas. Tai vienas iš patikimiausių metodų kalbai mokyti. Kalbos vertimo pratimai kurse pateikiami testų pavidalu. Reikia išversti angliškai parašytus žodžius į lietuvių kalbą.

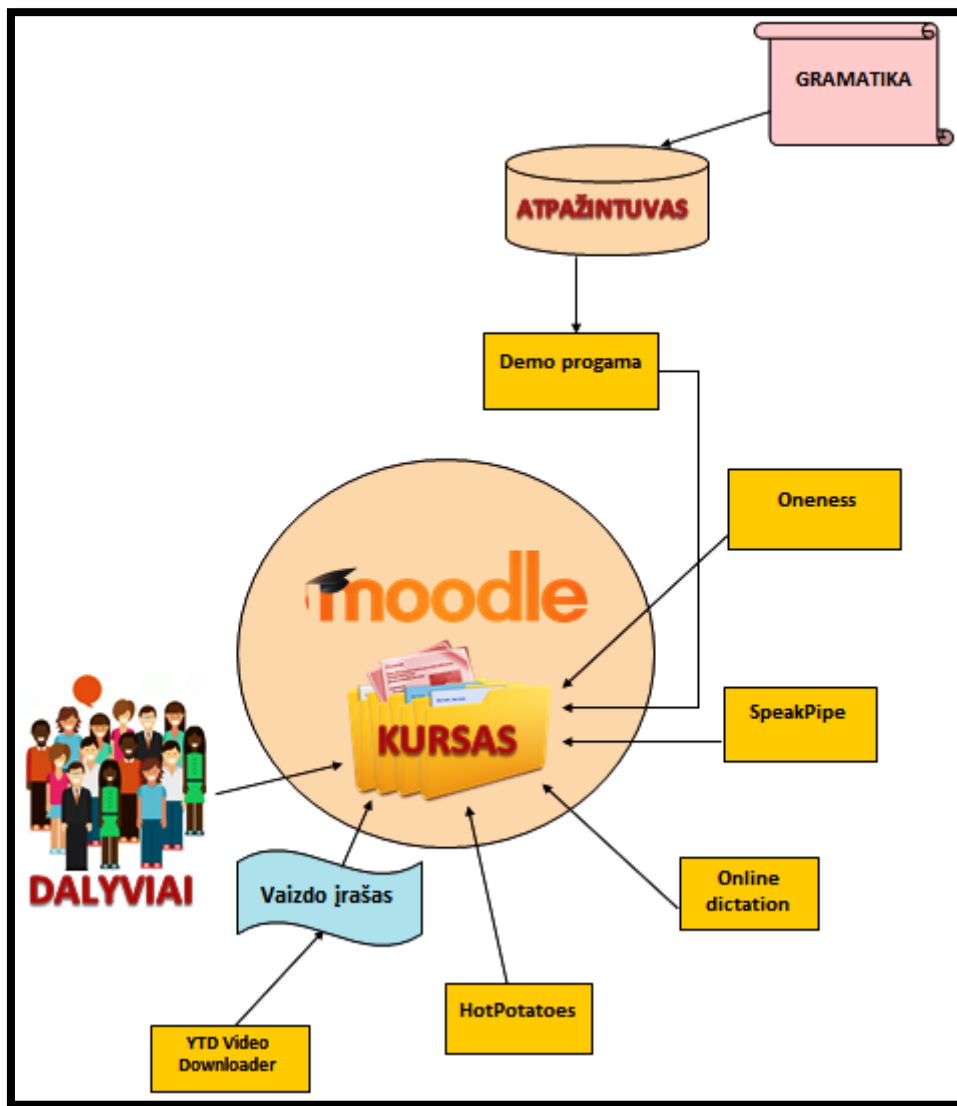
Gramatikos mokymas – tai teorinės medžiagos analizavimas. Svarbu įsigilinti į pateiktą mokymosi medžiagą, kad vėliau būtų galima įgytas žinias panaudoti praktiškai. Kiekvieną detalų analizavimą privalu patikrinti pateikiant mokymosi dalyviams užduotis: įrašyti reikiamą žodį į tuščią laukelį, parinkti sakinyje teisingą žodį. Taip mokomasi vartoti tam tikrą kalbos formą. Situacija lemia, kokia kalbos forma bus pasirinkta.

Kalbos tarties įgūdžiams tobulinti pateikiama užduočių, kuriose derinamas klausymas, tarimas, kalbėjimas. Tarties užduotys pateikiamos taip: besimokantysis privalo praktikuotis tarti žodžius ir frazes taip, kad kalbos atpažinimo įrankis teisingai juos atpažintų ir užrašytų virtualiame lape. Antruoju atveju, besimokantysis gali pats įvertinti savo tarimo įgūdžius diktuoamas frazes programai, kuri jas įrašo, vėliau jas galima perklaudyti. Trečiuoju atveju, nurodyti žodžiai ir frazės diktuojami atpažinimo programai, kuri kompiuterio ekrane parodo, ar teisingai jie buvo ištarti. Paskutiniuoju atveju, spontaniškos kalbos mokymui, besimokantiesiems pateikiamas vaizdo įrašas – daina su žodžiais. Dainavimas karaoke lavina kalbos spontaniškumą.

Derinant visas užduotis tarpusavyje, pateikiant jas besimokantiesiems, atitinkama tvarka gaunama mokymosi metodika, kurios rezultatas – įgytos žinios.

#### **4.5. Sistemos struktūra**

Lietuvių kalbos mokymosi sistema kuriama pagal metodiką, aprašytą 4.4. poskyryje. Sistemos struktūra pateikiama 11 pav.



**11 pav.** Virtuali lietuvių kalbos mokymo sistema

Sistemos branduolys yra virtuali mokymosi aplinka „Moodle“. Tai vieta, kurioje patalpintas šnekamosios lietuvių kalbos mokymo kursas.

Kursas yra orientuotas į užsienio studentus, jie sistemoje yra pažymėti kaip *dalyviai*.

Programa „YTD Video Downloader“ – tai programa, kurią naudojant galima parsisiųsti vaizdo įrašus iš tokių tinklalapių kaip: „Facebook“, „Youtube“, „Vevo“ ir kt. Naudojant programą „YTD Video Downloader“ parsųsti vaizdo įrašai naudojami kalbos mokymo tikslais lietuvių kalbos kurse, pvz. supažindinti besimokančiuosius su pagrindiniais faktais apie šalį, abėcėlės mokymo procese ir kt.

„Hot Potatoes“ – kompiuterinių programų rinkinys, kurį sudaro 5 skirtingos programos. Programų rinkinio paskirtis – kurti interaktyvius pratimus. Programų rinkinys sudarytas iš šių programų: „Jcloze“ – užduotims su praleistų žodžių įrašymu kurti, „Jcross“ – kryžiažodžiams sudarinėti, „Jmatch“ – užduotims, kuriose reikia išdėlioti atitikmenų poras, kurti, „Jmix“ – užduotims, kuriose reikia nustatyti žodžių eiliškumą sakinyje, kurti, „Jquiz“ – klausimynams kurti. Kryžiažodžiai kurse naudojami žinioms patikrinti.

„Online dictation“ – internetinis kalbos atpažinimo įrankis, atpažįstantis ir pateiktame virtualiame lape užrašantis visas ištartas balso komandas. Įrankis naudojamas kalbos tarties užduotims rengti, besimokančiųjų tarties įgūdžiams lavinti.

„SpeakPipe“ yra balso įrašymo įrankis, galimas rasti internete. Naudojant šį įrankį pateikiamos užduotys, kuriose reikia perklausyti ir įvertinti įdiktuotas frazes. Įrankio panaudojimas – kalbos tarčiai lavinti.

„Demo“ programa – kalbos atpažinimo programa, skirta kalbiniams lietuvių kalbos įgūdžiams lavinti. Pagrindinė programos idėja – ekrane išvysti teisingą rezultatą – padiktuotą balso komandą atitinkantį žodį ar frazę bei atpažintos komandos tikslumą nusakantį atpažinimo matą, pagal kurį galima spręsti, kaip tiksliai ji buvo atpažinta. Programoje išvedamas panašumo matas yra padaugintas iš 1000. Kuo arčiau 1000 ištartos frazės skaitinė reikšmė, tuo geresnis balso komandos atpažinimas.

„Oneness“ – kalbos mokymosi kursai, kuriuose gausu teorinės medžiagos, kuri yra paruošta Vilniaus universiteto lingvistų. Lietuvių kalbos mokymo kurse pateikiama nuorodų į „Oneness“ mokymosi medžiagą.

#### **4.6. Sistemos realizacija**

Kursui kurti pasirinkta mokymosi aplinka „Moodle“. „Moodle“ pasirinkta todėl, kad ją naudojami studentai. Atsižvelgiama ir į tai, kad kursą numatoma naudoti KTU universitete studijuojančių kitakalbių studentų lietuvių kalbiniams įgūdžiams formuoti. Kitakalbiams studentams studijuojant Lietuvos universitetuose privalu įgyti pradines lietuvių kalbos žinias. Tai galima padaryti virtualiame kurse, pateiktame paprastumu pasižyminčioje aplinkoje, naudojant netradicinius kalbos mokymosi metodus. Kursas „Moodle“ aplinkoje vadinasi „Lithuanian language for foreign students“, sutrumpintai „LLFFS16“.

Vaizdo įrašai, kurie yra naudojami kurse mokymosi tikslais yra atrenkami iš tinklalapio „Youtube“. Naudojant programą „YTD Video Downloader“ vaizdo įrašai yra parsisiunčiami į asmeninį kompiuterį, vėliau keliami į „Moodle“ parenkant funkciją „atidaryti tame pačiame lange“. Kitu atveju galima įdėti nuorodą į šaltinį, kuriame yra vaizdo įrašas, atidarant jį tame pačiame, arba kitame lange.

Mokymo kurse iš programų rinkinio „Hot Potatoes“ panaudojama JCross – kryžiažodžiams sudarinėti skirta programa. Kryžiažodžiai kurse naudojami įgytoms žinioms patikrinti. Sukūrus tam tikrą temą atitinkantį kryžiažodį šis patalpinamas kurse „Moodle“ aplinkoje, pritaikant tokio tipo užduotį prie mokymosi medžiagos.

Tarčiai lavinti kurse pateikiami specialūs įrankiai. Pirmiausia tai įrankis, laisvai prieinamas internete, „Online dictation“. Šis įrankis atpažįsta tariamus žodžius ir juos užrašo tinklalapyje pateiktame virtualiame popieriaus lape. Blogai ištartas žodis bus blogai užrašytas. Taigi atsižvelgiant



į priemonės pagrindinę funkciją, studentams kurse pateikiamas tekstas, kurį reikia diktuoti minėtajam įrankiui ir stebėti, ar viskas teisingai užrašoma. Bet norint kurse pateikti tokio tipo užduotį reikia gerai apgalvoti, kuriame etape teikti tokio tipo užduotį, suderinti ją su kitomis tartį lavinančiomis užduotimis. Trūkumas – šis įrankis gali būti naudojamas tik su „Google Chrome“ naršykle, kurioje yra integruotas kalbos atpažinimo variklis.

Antroji priemonė, skirta kalbos tarčiai lavinti – „Demo“ programa. Programa remiasi kalbos atpažinimo technologijomis, atpažįsta per mikrofoną jai padiktuotus žodžius ir ekrane išveda atpažintą žodį arba frazę – viskas priklauso nuo to, kaip buvo atpažinta balso komanda. Programai priskiriamas balso komandų sąrašas, kurias ji atpažins. Svarbu programos panaudojimą suderinti su konkrečiomis tarties užduotimis, kad programoje būtų užkrautas reikiamas žodžių sąrašas. Trūkumas – programa gali būti naudojama tik su „Mozilla Firefox“ naršykle, nes kitos naršyklės, pvz: „Google Chrome“ reikalauja saugumo sertifikato prieigai prie `get_user_data()` funkcijos, o „Internet Explorer“ minėtosios funkcijos apskritai nepalaiko.

Trečioji priemonė, skirta kalbos tarčiai lavinti, yra programa „SpeakPipe“. Tai laisvai internete prieinama programa su balso įrašymo ir atkūrimo funkcija. Programa kurse naudojama pritaikant konkrečioms užduotims, kuriuose studentas perskaito pateiktą tekstą jį įrašydamas į programą, perklauso įrašą ir bando vertinti savo tarties įgūdžius.

„Oneness“ lietuvių kalbos mokymo kursas – yra inovatyvus būdas mokytis lietuvių kalbos. Jame pateikiama lingvistų atrinkta lietuvių kalbos medžiaga, klausymo užduotys ir įdomūs testai. Siekiant išvengti persidengiančios informacijos apie kalbos gramatiką ir tam tikras temas, nuspręsta kalbos kurse naudoti nuorodas į „Oneness“ kurso medžiagą.

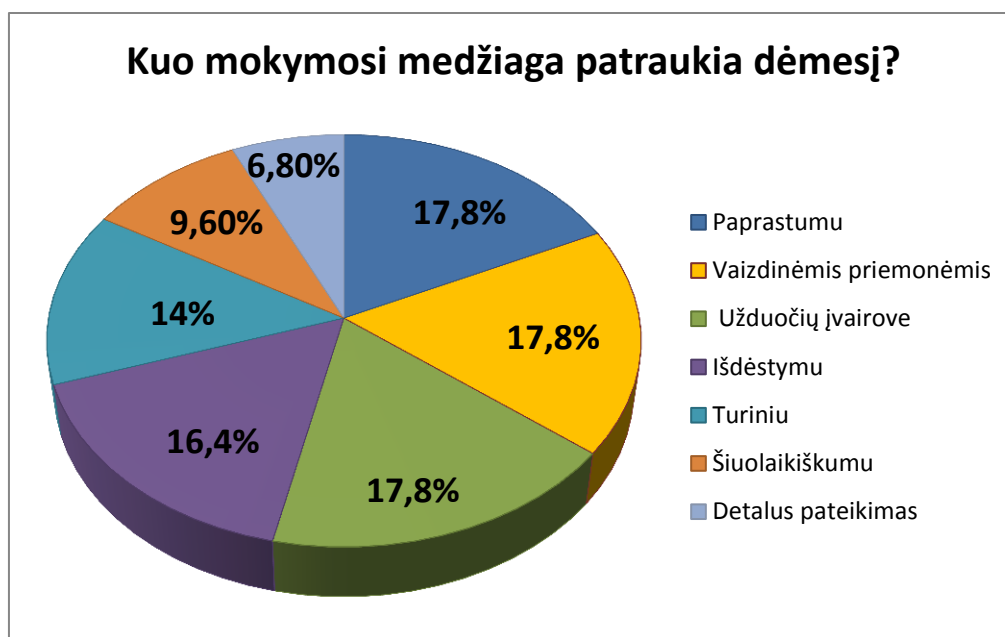
Apjungus visus šiuos komponentus kuriamas „Moodle“ kursas skirsis nuo jau esamų kursų ir mokymo priemonių, nes jame pateikiama mokymo užduočių, kurios remiasi kalbos atpažinimo technologijomis.

#### **4.7. Sistemos kokybės tyrimas**

Prieš kuriant virtualų kursą buvo atliekamos studentų ir ekspertų apklausos, siekiant išsiaiškinti lūkesčius kuriamam kalbos mokymosi kursui. Priimtiems sprendimams dėl kalbos mokymosi kurso kūrimo įvertinti, kalbos mokymosi kursas pateiktas Kauno technologijos universiteto Informatikos fakulteto bei Socialinių, humanitarinių mokslų ir menų fakulteto užsienio studentams. Apklausa išplatinta bendradarbiaujant su minėtųjų fakultetų dėstytojais. Apklausa sudaryta iš 15 klausimų.

Šiame poskyryje bus glaustai aptariami apklausos klausimai ir atsakymai. Detalią apklausos analizę galima rasti 21 priede. Atsakymus į apklausos klausimus pateikė 19 kursą išbandžiusių užsienio studentų, t.y. 11 (57,9%) vyrų ir 8 (42,1%) moterys. Didžioji dalis, net 10 (52,6%) studentų, teigia esantys iš Indijos, po 3 studentus (15,8%) iš Ispanijos ir Portugalijos, o likusieji 2 (10,5%) gimtąja šalimi nurodė esant Turkiją bei 1 (5,3%) Kiniją.

Pirmiausia studentų buvo paklausta, kuo mokymosi medžiaga patraukia dėmesį. Rezultatai pasiskirstė taip: paprastumu, vaizdinėmis priemonėmis bei užduočių įvairove (po 17,8%), medžiagos išdėstymas (16,4%), medžiagos turinys (13,7%), užduočių šiuolaikiškumas (9,6%) ir detalus informacijos pateikimas (6,8%). Nuomonių pasiskirstymas pavaizduotas 12 pav.



**12 pav.** Mokymosi medžiagai patrauklumą teikiantys aspektais

Dvikalbystę visame kurse studentai vertina taip pat skirtingai. Daugumai, net 12 (60%) tai yra priimtina, 5 (25%) tai nėra priimtina, o likusiems 3 (15%) nesvarbu, ar medžiaga ir jos išdėstymas yra teikiamas viena, ar dviem kalbomis.

Kalbinės medžiagos išdėstymas kurse, taip pat ir mokymosi procesą reguliuojantys nurodomieji žingsniai yra vienodai priimtini visiems 19-kai studentų (po 100%).

Toliau buvo paklausta apie kalbinių įgūdžių tobulinimą naudojant kalbos atpažinimo technologijas. Visi apklaustieji vienareikšmiškai sutiko, jog tai yra naudingas dalykas.

Taip pat studentų buvo paprašyta įvardinti labiausiai naudą teikiančius kurso dalykus. Daugiausiai, net 11, studentų teigia, kad viskas kurse jiems yra naudinga, 2-iems studentams – kalbinius įgūdžius lavinančios užduotys, o likusiems: medžiagos paprastumas bei virtualaus mokymosi teikiama galimybė mokytis pačių pasirinktu laiku, patogioje vietoje.

Nuomonės apie kurso trūkumus pasiskirstė įvairiai. 18 studentų atsakė, jog kurse nėra trūkumų, 1 studentas pasigedo galimybės pretenduoti į stipendijas.

Atlikta apklausa parodė, jog į KTU iš užsienio šalių atvykę studentai lietuvių kalbos mokymąsi virtualiu būdu laiko priimtiniu kalbos mokymosi būdu. Tikslingai mokymosi procese parinktos užduotys, jų paprastumas bei atitinkamų IT priemonių panaudojimas, studentams yra efektyvus žinių įgijimo būdas, nepaisant keleto trūkumų, kurie buvo įvardinti.

#### **4.8. Projektinės dalies išvados**

1. Siekiant išsiaiškinti lūkesčius virtualiam kalbos mokymosi kursui, atliktos ekspertų bei studentų apklausa. Remiantis apklausų rezultatais, parinktos mokymosi užduotys kursui kurti: kalbėjimo, klausymo, gramatikos suvokimo, rašymo. Užduotys teikiamos atsižvelgiant į studentų pageidautas temas: abėcėlės mokymas, savęs pristatymas.
2. Remiantis sukurta lietuvių kalbos mokymo metodika, parinktos priemonės, kuriomis kalbos mokymosi sistemoje įgyvendinamas mokymosi procesas. Naudojamos priemonės: vaizdo įrašai, kryžiažodžiai, teorinė medžiaga, interaktyvi mokymosi medžiaga, kalbos atpažinimo įrankiai.
3. Sukurta virtuali lietuvių kalbos mokymo sistema, kurios dalis yra kursas virtualioje mokymosi aplinkoje „Moodle“.
4. Atliktas virtualaus lietuvių kalbos kurso kokybės tyrimas parodė, kad kursas užsienio studentams yra įdomus, inovatyvus bei efektyvus būdas kalbai mokytis.

## IŠVADOS

1. Išanalizavus teorinius šaltinius galima teigti, jog visapusiškam kalbos išmokimui yra svarbu, kad būtų lavinami tiek kalbinės, tiek kultūrinės komunikacijos gebėjimai. Jie yra vienas nuo kito neatskiriami.
2. Išanalizavus internete surastas kalbos mokymosi priemonės paaiškėjo, jog jos nėra tinkamos visapusiškam lietuvių kalbos mokymuisi. Analizuojant priemones paaiškėjo, jog „SurfaceLanguages“ ir „Loecsen“ neturi į tarimo įgūdžius orientuotų užduočių. Priemonė „Duolingo“ atitinka visus kalbos mokymo aspektus, tačiau jos trūkumas – nėra galimybės mokytis lietuvių kalbos.
3. Projektuojant lietuvių kalbos mokymosi sistemą sukurti daugiadiktoriniai garsynai. Tai taisyklingai ištartų sudėtingų lietuviškų pavardžių ir vardų bei netaisyklingai tariamų tų pačių pavardžių ir vardų garsynai. Atlikus tyrimus su garsynais naudojant HTK programų paketą nustatyta aukščiausia atpažinimo kokybė esant fiksuotam būsenų skaičiui 14 ir 16-kai mišinių – 96,75 proc. Atlikus tyrimus ispanų kalbos atpažintuvu nustatytas kalbos atpažinimo tikslumas – 98,4 proc.
4. Remiantis HTK programų paketu ir ispanų kalbos atpažintuvu gautais tyrimų rezultatais buvo atrinktos komandos, kurios naudojamos sukurtoje balso atpažinimo programoje.
5. Išanalizavus studentų ir kalbos ekspertų apklausų rezultatus nuspręsta, kokios kalbos mokymosi užduotys bei priemonės bus naudojamos virtualiame kurse.
6. Atlikus virtualaus lietuvių kalbos kurso kokybės tyrimą paaiškėjo, kad kursas užsienio studentams yra įdomus, lavinantis visas kalbines sritis, efektyvus bei patogus dėl virtualaus pateikimo.

## LITERATŪRA

- [1] BALAVOČIUS, B. ir L. TELKSNYS. Garsynų duomenų modeliai ir programinės įrangos architektūros. *Lietuvių kalbos kompiuteriniai tyrimai (IX sekcija)*. Kaunas: Vytauto Didžiojo universitetas, 2003, IX–1-IX–9. Prieiga per internetą: [http://www.elibrary.lt/resursai/Konferencijos/KTU\\_01/IT\\_2003/Sekcija09.pdf](http://www.elibrary.lt/resursai/Konferencijos/KTU_01/IT_2003/Sekcija09.pdf)
- [2] BELEŽENTIS, A ir M. ŽALIMAITĖ. Ekspertinių vertinimų inovacijų plėtros veiksnių analizėje: Lietuvos įnovatyvių įmonių vertinimas. *Research papers*. Vilnius: Mykolo Romerio universitetas, 2011, vol.\_27, no 3, pp. 23-31. ISSN 1822-6760.
- [3] DUNN, D, Michael. Pro Microsoft Speech Server 2007: Developing Speech Enabled Applications with .NET. Jav. 2007. ISBN-13: 978-1-59059-902-0;
- [4] DŽEŽULSKIENĖ, Judita. Kalbos technologijų taikymo lietuvių kaip svetimai kalbai mokytis ypatumai. *Kalbų studijos*. 2014, 24, 106-112. Prieiga per: doi: <http://dx.doi.org/10.5755/j01.sal.0.24.7067>
- [5] FOSBERG, M. Why is speech recognition difficult? [interaktyvus]. 2007, 1-10 [žiūrėta 2015-04-16]. Prieiga per internetą: [http://www.speech.kth.se/~rolf/gslt\\_papers/MarkusForsberg.pdf](http://www.speech.kth.se/~rolf/gslt_papers/MarkusForsberg.pdf)
- [6] YOUNG, Steve ir k. t. (2006). The HTK Book (for HTK Version 3.1) [interaktyvus]. 2006 [žiūrėta 2015 01 10]. Prieiga per internetą <http://nesl.ee.ucla.edu/projects/ibadge/docs/ASR/htk/htkbook.pdf>
- [7] JURAFSKY, D. and J.H. MARTIN. Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall, Upper Saddle River, New Jersey 07458. Inc, 934 p.
- [8] KASPARAITIS, Pijus. Kirčiavimas. Transkribavimas. Lietuvių kalbos daiktavardžių ir būdvardžių kirčiavimas. Kompiuterinės lingvistikos paskaitų kospektai. 2005, 1-6. Prieiga per internetą: <http://www.mif.vu.lt/~pijus/CL/KircTr.pdf>
- [9] KAZLAUSKIENĖ, A., RAŠKINIS, G., VAIČIŪNAS, A. Automatinis lietuvių kalbos žodžių skiemonavimas, kirčiavimas, transkribavimas. Kaunas: Vytauto Didžiojo universitetas, 2010, pp. 2-107. ISBN 978-9955-12-630-0
- [10] LAURINČIUKAITĖ, S. Lietuvių šnekos atpažinimo akustinis modeliavimas: daktaro disertacija: technologijos mokslai, informatikos inžinerija 07T. Vilnius: Vilniaus Gedimino technikos universitetas, 2008, pp. 134. Prieiga per internetą: [www.mii.lt/files/mii\\_dis\\_08\\_laurinciukaite.pdf](http://www.mii.lt/files/mii_dis_08_laurinciukaite.pdf)
- [11] Lietuvos bendrojo lavinimo mokyklos bendrosios programos ir bendrojo išsilavinimo standartai. XI – XII klasės. 2012, Vilnius: Švietimo plėtotės centras.
- [12] LIUBINIENĖ, Vilmantė. Užsienio kalbos gebėjimų, integruotų su dalyko turiniu, lavinimas aukštojoje mokykloje: KTU atvejis. *Kalbų studijos*. 2010, 16, 101-105. Prieiga per internetą:

[http://www.kalbos.lt/zurnalai/16\\_numeris/16.pdf](http://www.kalbos.lt/zurnalai/16_numeris/16.pdf)

[13] MASKELIŪNAS, R. Lietuviškų balso komandų atpažinimas daugybinių transkripcijų pagrindu: daktaro disertacijos santrauka 07T. Kaunas: Kauno technologijos universitetas, 2009, pp. 34. Prieiga per internetą: [http://en.ktu.lt/sites/default/files/2009-12-14%20Maskeliunas\\_santr.pdf](http://en.ktu.lt/sites/default/files/2009-12-14%20Maskeliunas_santr.pdf)

[14] MAUZIENĖ, L. Leksikos mokymo aktualijos. *Filologija*. Vilnius: Vilniaus Gedimino technikos universitetas, 2005, 60-64. Prieiga per internetą: <http://archive.minfolit.lt/arch/3501/3584.pdf>

[15] MIKELIONIENĖ, Jurgita. Anototas sakininės kalbos tekstynas – intermedialiomis technologijomis paremta kalbų mokymo priemonė. *Kalbų studijos*. 2009, 14, pp. 39–43. Prieiga per internetą: [http://www.kalbos.lt/zurnalai/14\\_numeris/07.pdf](http://www.kalbos.lt/zurnalai/14_numeris/07.pdf)

[16] *Pamokykite kolegas lietuvių kalbos*. 2016 [interaktyvus]. Prieiga per internetą: <http://www.lietuva.lu/?q=node/404>. [žiūrėta 2016-04-30].

[17] PRANEVIČIUS, Henrikas ir kt. *Agentinių sistemų modeliai: mokomoji knyga*. Vilnius: Mokslo aidai, 2008. ISBN 978-9955-591-55-9

[18] RAŠKINIS, A., RAŠKINIS G., KAZLAUSKIENĖ A. VDU bendrinės lietuvių šnekos universalus anototas garsynas. *Lietuvių kalbos kompiuteriniai tyrimai (IX sekcija)*. Kaunas: Vytauto Didžiojo universitetas, 2003. Prieiga per internetą: [http://www.ktu.lt/lt/apie\\_renginius/konferencijos/2006/k6\\_02/it2003/sekcija09.pdf](http://www.ktu.lt/lt/apie_renginius/konferencijos/2006/k6_02/it2003/sekcija09.pdf)

[19] RATKEVIČIUS, K. Hibridinė atpažinimo technologija balso sąsajai (Infobalsas): ataskaita. Kaunas: Kauno technologijos universitetas, 2013. pp.50.

[20] RINGELIENĖ, Živilė ir Mark FILIPOVIČ. Žodžių atpažinimo, grįsto paslėptaisiais Markovo modeliais, vizualizavimo ir analizės programinė įranga. *Informacijos mokslai*. 2011, 63-72. ISSN 1392-0561. Prieiga per internetą: <http://www.vu.lt/leidyba/dokumentai/zurnalai/INFORMACIJOS%20MOKSLAI/Informacijos%20mokslai%202011%2056/63-72.pdf>

[21] RUDŽIONIS, A., RATKEVIČIUS, K., RUDŽIONIS, V., KASPARAITIS, P., ŠALNA, P. Automatinis lietuvių kalbos žodžių skiemonavimas, kirčiavimas, transkribavimas. Kaunas, Vilnius: Kauno technologijos universitetas, Vilniaus universitetas, Teismo ekspertizės centras. 2001, pp. 84. Prieiga per internetą: [http://www.likit.lt/all/balso\\_tech/balsotech.zip](http://www.likit.lt/all/balso_tech/balsotech.zip)

[22] RUDŽIONIS, Vytautas., RAŠKINIS, Gailius., MASKELIŪNAS, Rytis., RUDŽIONIS, Algimantas., RATKEVIČIUS, Kastytis., BARTIŠIŪTĖ, Gintarė. Web services based hybrid recognizer of Lithuanian voice commands. *Elektronika ir elektrotechnika*. 2014, 20 (9), 50-53. Prieiga per internetą: <http://www.eejournal.ktu.lt/index.php/elt/article/viewFile/8713/4289>

[23] TAMULEVIČIUS, G. Pavienių žodžių atpažinimo sistemų kūrimas: daktaro disertacija: technologijos mokslai, informatikos inžinerija 07T. Vilnius: Vilniaus Gedimino technikos

universitetas, 2008, pp.124. Prieiga per internetą:

[http://donelaitis.vdu.lt/disertacijos/Tamulevicius\\_2008.pdf](http://donelaitis.vdu.lt/disertacijos/Tamulevicius_2008.pdf)

[24] *Tinklapių Infobitas Info informacija* [interaktyvus]. 2011. [žiūrėta 2015-03-14]. Prieiga per internetą:

[http://info.infobitas.lt/index.php?option=com\\_content&view=article&id=74:balso-atpazinimo-sistemas&catid=116&Itemid=481](http://info.infobitas.lt/index.php?option=com_content&view=article&id=74:balso-atpazinimo-sistemas&catid=116&Itemid=481)

[25] VAIČIŪNAS, A. Lietuvių kalbos statistinių modelių ir jų taikymo šnekos atpažinimui tyrimas, kai naudojami labai dideli žodynai: daktaro disertacija: fiziniai mokslai, informatika 09P. Kaunas: Vytauto Didžiojo universitetas, 2006, pp. 139. Prieiga per internetą:

[http://donelaitis.vdu.lt/disertacijos/Vaiciunas\\_2006.pdf](http://donelaitis.vdu.lt/disertacijos/Vaiciunas_2006.pdf)

[26] ŽUKIENĖ, Regina. Informacinių technologijų panaudojimas mokant VGTU studentus užsieniečius lietuvių kalbos kaip svetimos. 2011, pp. 108-116. Prieigas per: doi: 10.3846/cpe.2011.12

## 1 PRIEDAS LIETUVIŠKŲ PAVARDŽIŲ IR VARDŲ GARSYNAS

<b>Nr.</b>	<b>Balso komanda</b>	<b>Katalogas</b>
1.	Adomavičiūtė Inga	ADOMAVICIUTE
2.	Aleksandravičiūtė Ramutė	ALEKSANDRAVICIUTE
3.	Ališauskienė Jadvyga	ALISAUSKIENE
4.	Bulachienė Danguolė	BULACHIENE
5.	Čizaitė Radvilė	CIZAITE
6.	Danilkevičius Rytis	DANILKEVICIUS
7.	Eidžiūnas Aurimas	EIDZIUNAS
8.	Grigaravičienė Jūratė	GRIGARAVICIENE
9.	Ignatavičius Gintaras	IGNATAVICIUS
10.	Ivancevičiūtė Erika	IVANCEVICIUTE



**2 PRIEDAS PAVARDŽIŲ IR VARDŲ ATPAŽINIMO TIKSLUMO PRIKLAUSOMYBĖ  
NUO BŪSENŲ SKAIČIAUS**

<b>Atpažinimo tikslumo priklausomybė nuo būsenų skaičius, %</b>					
<b>Vardas ir pavardė</b>	<b>Būsenų skaičius atitinka raidžių sk.</b>				
	<b>+0 būsenų</b>	<b>+2 būsenos</b>	<b>+4 būsenos</b>	<b>+5 būsenos</b>	<b>+6 busenos</b>
<b>Adomavičiūtė Inga</b>	85	50	47,5	50	87,5
<b>Aleksandravičiūtė Ramutė</b>	67	100	100	100	100
<b>Ališauskienė Jadvyga</b>	57	0	100	100	40
<b>Bulachienė Danguolė</b>	50	75	100	97,5	97,5
<b>Čizaitė Radvilė</b>	40	100	12,5	2,5	85
<b>Danilkevičius Rytis</b>	65	52	100	100	100
<b>Eidžiūnas Aurimas</b>	97	80	100	72,5	100
<b>Grigaravičienė Jūratė</b>	50	35	57,5	50	50
<b>Ignatavičius Gintaras</b>	50	100	100	50	15
<b>Ivancevičiūtė Ramutė</b>	52	82	87,5	50	100
<b>Vidurkis, %</b>	<b>61,5</b>	<b>67,5</b>	<b>80,5</b>	<b>67,25</b>	<b>77,5</b>

**3 PRIEDAS ATPAŽINIMO TIKLSUMO PRIKLAUSOMYBĖ NUO GAUSO SKIRSTINIŲ SKAIČIAUS IR +4 BŪSENŲ**

<b>Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus, %</b>		
<b>Vardas ir pavardė</b>	<b>Būsenų skaičius atitinka raidžių sk. + 4</b>	
	<b>2 mišiniai</b>	<b>6 mišiniai</b>
<b>Adomavičiūtė Inga</b>	70	55
<b>Aleksandravičiūtė Ramutė</b>	50	60
<b>Ališauskienė Jadvyga</b>	95	90
<b>Bulachienė Danguolė</b>	100	80
<b>Čizaitė Radvilė</b>	32,5	75
<b>Danilkevičius Rytis</b>	50	47,5
<b>Eidžiūnas Aurimas</b>	100	60
<b>Grigaravičienė Jūratė</b>	55	52,5
<b>Ignatavičius Gintaras</b>	52,5	60
<b>Ivancevičiūtė Ramutė</b>	50	52,5
<b>Vidurkis, %</b>	<b>65,5</b>	<b>63,25</b>

**4 PRIEDAS ATPAŽINIMO TIKSLUMO PRIKLAUSOMYBĖ NUO GAUSO SKIRSTINIŲ SKAIČIAUS IR FIKSUOTO 14 BŪSENŲ SKAIČIAUS**

<b>Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus, %</b>		
<b>Vardas ir pavardė</b>	<b>Fiksuotas būsenų skaičius 14</b>	
	<b>4 mišiniai</b>	<b>16 mišinių</b>
<b>Adomavičiūtė Inga</b>	55	100
<b>Aleksandravičiūtė Ramutė</b>	60	100
<b>Ališauskienė Jadvyga</b>	37,5	100
<b>Bulachienė Danguolė</b>	75	100
<b>Čizaitė Radvilė</b>	72,5	100
<b>Danilkevičius Rytis</b>	50	100
<b>Eidžiūnas Aurimas</b>	82,5	100
<b>Grigaravičienė Jūrate</b>	62,5	67,5
<b>Ignatavičius Gintaras</b>	72,5	100
<b>Ivancevičiūtė Ramutė</b>	50	100
<b>Vidurkis, %</b>	<b>61,75</b>	<b>96,75</b>

**5 PRIEDAS ATPAŽINIMO TIKSLUMO PRIKLAUSOMYBĖ NUO GAUSO SKIRSTINIŲ SKAIČIAUS IR FIKSUOTO 16 BŪSENŲ SKAIČIAUS**

<b>Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus, %</b>		
<b>Vardas ir pavardė</b>	<b>Fiksuotas būsenų skaičius 16</b>	
	<b>4 mišiniai</b>	<b>16 mišinių</b>
<b>Adomavičiūtė Inga</b>	57,5	100
<b>Aleksandravičiūtė Ramutė</b>	95	100
<b>Ališauskienė Jadvyga</b>	75	100
<b>Bulachienė Danguolė</b>	77,5	100
<b>Čizaitė Radvilė</b>	55	100
<b>Danilkevičius Rytis</b>	32,5	100
<b>Eidžiūnas Aurimas</b>	92,5	100
<b>Grigaravičienė Jūratė</b>	95	67,5
<b>Ignatavičius Gintaras</b>	65	100
<b>Ivancevičiūtė Ramutė</b>	55	100
<b>Vidurkis, %</b>	<b>70</b>	<b>96,75</b>

**6 PRIEDAS ATPAŽINIMO TIKSLUMO PRIKLAUSOMYBĖ NUO GAUSO SKIRSTINIŲ SKAIČIAUS IR FIKSUOTO 17 IR 21 BŪSENŲ SKAIČIAUS**

<b>Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus, %</b>		
<b>Vardas ir pavardė</b>	<b>Fiksuotas būsenų skaičius</b>	
	<b>17 ir 16 mišinių</b>	<b>21 ir 16 mišinių</b>
<b>Adomavičiūtė Inga</b>	82,5	92,5
<b>Aleksandravičiūtė Ramutė</b>	97,5	100
<b>Ališauskienė Jadvyga</b>	95	57,5
<b>Bulachienė Danguolė</b>	100	100
<b>Čizaitė Radvilė</b>	100	100
<b>Danilkevičius Rytis</b>	100	52,5
<b>Eidžiūnas Aurimas</b>	100	100
<b>Grigaravičienė Jūratė</b>	72,5	52,5
<b>Ignatavičius Gintaras</b>	100	97,5
<b>Ivancevičiūtė Ramutė</b>	77,5	57,5
<b>Vidurkis, %</b>	<b>92,5</b>	<b>81</b>

**7 PRIEDAS NETAISYKLINGAI IŠTARTŲ PAVARDŽIŲ IR VARDŲ ATPAŽINIMO  
TIKLŲ PARIKLAUSOMYBĖ NUO GAUSO SKIRSTINIŲ IR FIKSUOTO BŪSENŲ  
SKAIČIAUS**

<b>Atpažinimo tikslumo priklausomybė nuo Gauso skirstinių skaičiaus, %</b>		
<b>Vardas ir pavardė</b>	<b>Fiksuotas būsenų skaičius</b>	
	<b>14 ir 16 mišinių</b>	<b>16 ir 16 mišinių</b>
<b>Adomavičiūtė Inga</b>	64	44
<b>Aleksandravičiūtė Ramutė</b>	54	38
<b>Ališauskienė Jadvyga</b>	48	32
<b>Bulachienė Danguolė</b>	26	34
<b>Čizaitė Radvilė</b>	0	0
<b>Danilkevičius Rytis</b>	12	18
<b>Eidžiūnas Aurimas</b>	96	94
<b>Grigaravičienė Jūratė</b>	30	54
<b>Ignatavičius Gintaras</b>	78	52
<b>Ivancevičiūtė Ramutė</b>	62	74
<b>Vidurkis, %</b>	<b>47</b>	<b>44</b>

## 8 PRIEDAS BALSŲ KOMANDOS GRAMATIKAI NR. 1

1. Šviečia saulė
2. Aš esu Marta
3. Vaikinas
4. Mergina
5. Kokia tavo tautybė?
6. Kalbėti
7. Kalbu lietuviškai
8. Kur mokaisi?
9. Mokausi universitete
10. Auditorija
11. Aš informatikas.
12. Kiek tau metų?
13. Man dvidešimt du.
14. Namas
15. Butas
16. Ponas
17. Kur eini?
18. Einu į paskaitą.
19. Laisvalaikis.
20. Man patinka krepšinis
21. Pavardė
22. Aukštas
23. Kursas
24. Laimingas
25. Liūdnas
26. Darbo laikas
27. Troleibusas
28. Dangus
29. Užsienio studentas
30. Protingas

## 9 PRIEDAS GRAMATIKA NR. 1

```
<grammar xmlns:sapi="http://schemas.microsoft.com/Speech/2002/06/SRGSExtensions" xml:lang="es-ES" tag-  
format="semantics-ms/1.0" version="1.0" mode="voice" xmlns="http://www.w3.org/2001/06/grammar"  
sapi:alphabet="x-microsoft-ups">
```

```
  <rule id="Rule" scope="public">  
    <one-of>
```

```
</item>
```

```
  <one-of>  
    <item>sviecia saule</item>  
    <item>svjiecja saule</item>  
  </one-of>  
  <tag>$. _value = "Sviecia saule"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>ash esju marta</item>  
    <item>ash jesju mjarta</item>  
  </one-of>  
  <tag>$. _value = "As esu Marta"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>vajkynjas</item>  
    <item>vajkinjas</item>  
  </one-of>  
  <tag>$. _value = "Vaikinas"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>mergjynja</item>  
    <item>merginja</item>  
  </one-of>  
  <tag>$. _value = "Mergina"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>kokja tavjo tjautybje</item>  
    <item>kjokja tjavjo tjautybje</item>  
  </one-of>  
  <tag>$. _value = "Kokia tavo tautybe?"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>kalbeti</item>  
    <item>kalbjeti</item>  
  </one-of>  
  <tag>$. _value = "Kalbeti"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>kalbju ljetuvjyskai</item>  
    <item>kalbju ljatuvjyshkai</item>  
  </one-of>  
  <tag>$. _value = "Kalbu lietuviskai"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>kjur mokjaisi</item>  
    <item>kur mokjaisy</item>  
  </one-of>  
  <tag>$. _value = "Kur mokaisi?"</tag>
```

```
</item>
```



<item>		<one-of> <item>mokjausi unjyversytetje</item> <item>mokaujsy unjyversytjetje</item> </one-of> <tag>\$. _value = "Mokausi universitete"</tag>
	</item>	
<item>		<one-of> <item>audortija</item> <item>audytoryja</item> </one-of> <tag>\$. _value = "Auditorija"</tag>
	</item>	
<item>		<one-of> <item>ash informjatyjkas</item> <item>ash ynformatjykas</item> </one-of> <tag>\$. _value = "As informatikas"</tag>
	</item>	
<item>		<one-of> <item>kjiek tau metju</item> <item>kiek tjau metju</item> </one-of> <tag>\$. _value = "Kiek tau metu?"</tag>
	</item>	
<item>		<one-of> <item>mjan dvydjesymtj dju</item> <item>man dvidjesymt du</item> </one-of> <tag>\$. _value = "Man dvidesimt du"</tag>
	</item>	
<item>		<one-of> <item>namjas</item> <item>njamjas</item> </one-of> <tag>\$. _value = "Namas"</tag>
	</item>	
<item>		<one-of> <item>bjutjas</item> <item>butjas</item> </one-of> <tag>\$. _value = "Butas"</tag>
	</item>	
<item>		<one-of> <item>pjonas</item> <item>ponjas</item> </one-of> <tag>\$. _value = "Ponas"</tag>
	</item>	
<item>		<one-of> <item>kjur einji</item> <item>kur enji</item> </one-of> <tag>\$. _value = "Kur eini?"</tag>
	</item>	

<item>		<one-of> <item>einju y paskajtja</item> <item>ejnju y pjaskajtja</item> </one-of> <tag>\$. _value = "Einu i paskaita"</tag>
	</item>	
<item>		<one-of> <item>laisvalaikis</item> <item>lajsvajkis</item> </one-of> <tag>\$. _value = "Laisvalaikis"</tag>
	</item>	
<item>		<one-of> <item>mjan patjynka krepshynjis</item> <item>majn patynkja krepshynjis</item> </one-of> <tag>\$. _value = "Man patinka krepsinis"</tag>
	</item>	
<item>		<one-of> <item>pavarde</item> <item>pavardje</item> </one-of> <tag>\$. _value = "Pavarde"</tag>
	</item>	
<item>		<one-of> <item>aukshtjas</item> <item>aukshtas</item> </one-of> <tag>\$. _value = "Aukstas"</tag>
	</item>	
<item>		<one-of> <item>kjursas</item> <item>kursas</item> </one-of> <tag>\$. _value = "Kursas"</tag>
	</item>	
<item>		<one-of> <item>lajmyngas</item> <item>laimingjas</item> </one-of> <tag>\$. _value = "Laimingas"</tag>
	</item>	
<item>		<one-of> <item>ljudnas</item> <item>ljiudnas</item> </one-of> <tag>\$. _value = "Liudnas"</tag>
	</item>	
<item>		<one-of> <item>darbo laikas</item> <item>darbjo lajkas</item> </one-of> <tag>\$. _value = "Darbo laikas"</tag>
	</item>	

```

<item>
    <one-of>
    <item>trolejbusjas</item>
    <item>troleibusas</item>
    </one-of>
    <tag>$_value = "Troleibusas"</tag>
</item>
<item>
    <one-of>
    <item>dangus</item>
    <item>dangjus</item>
    </one-of>
    <tag>$_value = "Dangus"</tag>
</item>
<item>
    <one-of>
    <item>uzsienio stjudenjtjas</item>
    <item>uzsienio studentas</item>
    </one-of>
    <tag>$_value = "Uzsienio studentas"</tag>
</item>
<item>
    <one-of>
    <item>protyngjas</item>
    <item>protingas</item>
    </one-of>
    <tag>$_value = "Protingas"</tag>
</item>
</one-of>
</rule>
</grammar>

```

## 10 PRIEDAS BALSO KOMANDOS GRAMATIKAI NR.2

1. Upė
2. Ežeras
3. Žuvis
4. Plaukioju
5. Vaikštinėti
6. Vėjas
7. Gražūs vaizdai
8. Peizažas
9. Nuvykti
10. Temperatūra
11. Saulėta diena
12. Oras
13. Draugai
14. Ilsėtis
15. Man smagu
16. Atostogos
17. Šiandien
18. Prie jūros
19. Mėgautis
20. Grybauju
21. Uogauti
22. Geras laimikis
23. Puikus žvejys
24. Važinėti dviračiu
25. Skinu gėles

## 11 PRIEDAS GRAMATIKA NR. 2

```
<grammar xmlns:sapi="http://schemas.microsoft.com/Speech/2002/06/SRGSExtensions" xml:lang="es-ES" tag-  
format="semantics-ms/1.0" version="1.0" mode="voice" xmlns="http://www.w3.org/2001/06/grammar"  
sapi:alphabet="x-microsoft-ups">
```

```
  <rule id="Rule" scope="public">  
    <one-of>
```

```
</item>
```

```
  <one-of>  
    <item>upe</item>  
    <item>upje</item>  
  </one-of>  
<tag>$. _value = "Upe"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>ezeras</item>  
    <item>ezerjas</item>  
  </one-of>  
<tag>$. _value = "Ezeras"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>zuvys</item>  
    <item>zjuvys</item>  
  </one-of>  
<tag>$. _value = "Zuvys"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>plaukioju</item>  
    <item>pljaukioju</item>  
  </one-of>  
<tag>$. _value = "Plaukioju"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>vaikstineti</item>  
    <item>vajkstinjeti</item>  
  </one-of>  
<tag>$. _value = "Vaikstineti"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>vejas</item>  
  </one-of>  
<tag>$. _value = "Vejas"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>grazus vaizdai</item>  
    <item>grazjus vajzdaj</item>  
  </one-of>  
<tag>$. _value = "Grazus vaizdai"</tag>
```

```
</item>
```

```
<item>
```

```
  <one-of>  
    <item>peizazas</item>  
    <item>pejazazas</item>  
  </one-of>  
<tag>$. _value = "Peizazas"</tag>
```

```
</item>
```

```
<item>
```

		<pre> &lt;one-of&gt; &lt;item&gt;nuvykti&lt;/item&gt; &lt;item&gt;nuvyktji&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Nuvykti" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;tempjeratjura&lt;/item&gt; &lt;item&gt;temperatura&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Temperatura" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;sauleta diena&lt;/item&gt; &lt;item&gt;sauletja djiena&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Sauleta diena" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;oras&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Oras" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;draugai&lt;/item&gt; &lt;item&gt;draugaj&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Draugai" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;ilsjetis&lt;/item&gt; &lt;item&gt;ilsetis&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Ilsetis" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;man smagu&lt;/item&gt; &lt;item&gt;man smagju&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Man smagu" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;atostogos&lt;/item&gt; &lt;item&gt;atostjogos&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Atostogos" &lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;sjandien&lt;/item&gt; &lt;item&gt;siandien&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Siandien" &lt;/tag&gt; </pre>
<item>	</item>	

```

</item>
<item>
  <one-of>
    <item>prjie juros</item>
    <item>prie juros</item>
  </one-of>
  <tag>$. _value = "Prie juros" </tag>
</item>
<item>
  <one-of>
    <item>megautis</item>
    <item>mjegautjis</item>
  </one-of>
  <tag>$. _value = "Megautis" </tag>
</item>
<item>
  <one-of>
    <item>grybauju</item>
  </one-of>
  <tag>$. _value = "Grybauju" </tag>
</item>
<item>
  <one-of>
    <item>uogauti</item>
    <item>iuogjauti</item>
  </one-of>
  <tag>$. _value = "Uogauti" </tag>
</item>
<item>
  <one-of>
    <item>gjeras lajmikis</item>
    <item>geras laimikis</item>
  </one-of>
  <tag>$. _value = "Geras laimikis" </tag>
</item>
<item>
  <one-of>
    <item>pujkus zvejys</item>
    <item>puikus zvejys</item>
  </one-of>
  <tag>$. _value = "Puikus zvejys" </tag>
</item>
<item>
  <one-of>
    <item>vazinetis dviraciu</item>
    <item>vazjinetjis dviracju</item>
  </one-of>
  <tag>$. _value = "Vazinetis dviraciu" </tag>
</item>
<item>
  <one-of>
    <item>skinju gjeles</item>
    <item>skinu geles</item>
  </one-of>
  <tag>$. _value = "Skinu geles" </tag>
</item>
</one-of>
</rule>
</grammar>

```

## **12 PRIEDAS BALSO KOMANDOS GRAMATIKAI NR.3**

1. Einu į paskaitą
2. Mokyti
3. Skaitau knygą
4. Paskaitėlė
5. Išmintingas
6. Studijuoju
7. Noriu žinoti
8. Ilgai trunka
9. Vyksta
10. Važiuoju troleibusu
11. Bendrabutis
12. Grįžtu namo
13. Išsilavinimas
14. Kavos pertrauka
15. Kur yra biblioteka?
16. Knygų sąrašas
17. Tekstai
18. Namų darbai
19. Suprantu
20. Atsiskaitymas



### 13 PRIEDAS GRAMATIKA NR. 3

```
<grammar xmlns:sapi="http://schemas.microsoft.com/Speech/2002/06/SRGSExtensions" xml:lang="es-ES" tag-format="semantics-ms/1.0" version="1.0" mode="voice" xmlns="http://www.w3.org/2001/06/grammar" sapi:alphabet="x-microsoft-ups">
```

```
  <rule id="Rule" scope="public">
    <one-of>
```

```
</item>
```

```
  <one-of>
    <item>eiti y paskaita</item>
    <item>ejti y paskajta</item>
  </one-of>
  <tag>$. _value = "Eiti i paskaita"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>mokytis</item>
    <item>mokytjis</item>
  </one-of>
  <tag>$. _value = "Mokytis"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>skaitau knyga</item>
    <item>skajtau knyga</item>
  </one-of>
  <tag>$. _value = "Skaitau knyga"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>paskaitale</item>
    <item>paskajtjele</item>
  </one-of>
  <tag>$. _value = "Paskaitale"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>ismintingas</item>
    <item>ismintjingas</item>
  </one-of>
  <tag>$. _value = "Ismintingas"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>studijuoju</item>
  </one-of>
  <tag>$. _value = "Studijuoju"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>njoriu zinjoti</item>
    <item>noriu zinoti</item>
  </one-of>
  <tag>$. _value = "Noriu zinoti"</tag>
```

```
</item>
```

```
</item>
```

```
  <one-of>
    <item>ilgaj trunka</item>
    <item>ilgai trunka</item>
  </one-of>
  <tag>$. _value = "Ilgai trunka"</tag>
```

```
</item>
```

```
</item>
```

		<pre> &lt;one-of&gt; &lt;item&gt;vyksta&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Vyksta"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;vaziuoju troleibusu&lt;/item&gt; &lt;item&gt;vazjuoju trolejbusu&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Vaziuoju troleibusu"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;bendrabutis&lt;/item&gt; &lt;item&gt;bendrajbutjis&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Bendrabutis"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;gryztju namo&lt;/item&gt; &lt;item&gt;griztu namo&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Griztu namo"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;issilavinimas&lt;/item&gt; &lt;item&gt;issilavjinimas&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Issilavinimas"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;kavjos pertrauka&lt;/item&gt; &lt;item&gt;kavos pertrauka&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Kavos pertrauka"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;kjur yra bibliotekja&lt;/item&gt; &lt;item&gt;kur yra biblioteka&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Kur yra biblioteka?"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;knygju sarasjas&lt;/item&gt; &lt;item&gt;knygu sarasas&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Knygu sarasas"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; &lt;item&gt;tekstjai&lt;/item&gt; &lt;item&gt;tekstai&lt;/item&gt; &lt;/one-of&gt; &lt;tag&gt;\$. _value = "Tekstai"&lt;/tag&gt; </pre>
<item>	</item>	
		<pre> &lt;one-of&gt; </pre>



## 14 PRIEDAS LIETUVOS STUDENTŲ APKLAUSOS KLAUSIMAI

1. Lytis.
2. Amžius.
3. Mokymosi įstaiga.
4. Aš esu: dirbantis studentas / nedirbantis studentas.
5. Kokią kalbą norėtumėte išmokti artimiausioje ateityje?
6. Koks kalbų mokymosi metodas yra efektyviausias?
7. Naujai išmoktos kalbos vartojimo sritys.
8. Kokio tipo kalbos įgūdžių formavimas yra svarbiausias kalbos mokyme?
9. Kokių žinių norėtumėte įgyti mokydami kalbos?
10. Efektyviausios kalbos mokymosi užduotys.
11. Kokia virtualaus kalbos mokymo metodika Jūsų manymu yra efektyviausia?
12. Dėl kokių priežasčių rinktumėte kalbos mokymąsi virtualiu būdu?
13. Kokioje virtualioje erdvėje norėtumėte gauti mokymosi medžiagą/informaciją?
14. Ar kalbos mokymasis virtualiu būdu būtų efektyvus?

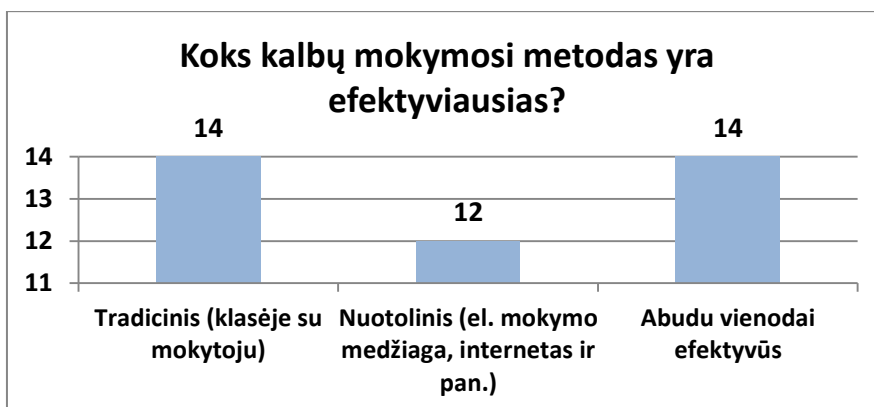
## 15 PRIEDAS LIETUVOS STUDENTŲ POREIKIŲ ANALIZĖ (PILNA)

Siekiant ištirti studentų poreikį studijuoti svetimą užsienio kalbą nuotoliniu būdu, buvo sugeneruota apklausa [apklausa.lt](http://apklausa.lt) tinklalapyje, kuria buvo apklausti 40 studentų iš skirtingų universitetų. Apklauskos anketą sudaro 14 klausimų, iš kurių keletas bendro pobūdžio (lytis, amžius, mokymosi įstaiga ir pan.), o likusieji susiję su informacijos rinkimu apie užsienio kalbos mokymą nuotoliniu būdu. Apklauskos platinimas vyko naudojantis socialiniu tinklalapiu „Facebook“.

Anketos klausimus galima išskirti į tris skirtingus tipus: bendro pobūdžio, susijusius su nuotoliniu mokymu ir su kalbos mokymu/pačios kalbos ypatybėmis. Pirmiausia bus aptariami bendrojo pobūdžio klausimai. Taigi atsižvelgiant į surinktus duomenis, apklausoje savo nuomonę pareiškė 18 moterų (45%) ir 22 vyrai (55%). 31-ienas apklaustasis teigė esantys asmenys iki 25 metų, iki 30 metų skalę papildė 6 asmenys, o likusieji 3 yra vyresni nei 30 metų asmenys. Tam, kad būtų aiškūs studentų poreikiai skirtinguose universitetuose, apklaustųjų buvo paprašyta nurodyti, kokioje įstaigoje šiuo metu studijuoja. Dauguma apklaustųjų (19 stud. – 47,5%) nurodė esantys iš KTU, 10 (25%) nurodė savo mokslo įstaigą esant VDU. Likusieji pasiskirstė taip: po 5 asmenys (12,5%) iš VGTU ir VU, ir 1 (2,5%) nurodė, jog yra LSU studentas. Kadangi minėtoji apklausa yra apie nuotolinį užsienio kalbos mokymą, todėl buvo svarbu sužinoti, kiek iš pasisakiusių yra dirbantys studentai, o kiek – bedarbiai, nes, tikėtina, jog užsienio kalbos kursus nuotoliniu būdu labiau bus linkę rinktis asmenys, turintys darbą ir siekiantys viską suderinti, kuomet bedarbiai (turintys daugiau laisvo laiko), turės didesnę galimybę rinktis tradiciniu mokymusi paremtus kursus/studijas. Ši iškelta hipotezė pasitvirtino, nes 28 atsakiusieji į klausimus yra dirbantys studentai, tai sudaro 70% visų apklaustųjų, o 30% – 12 apklaustųjų pažymėjo esantys nedirbantys studentai.

Su kalba susijusių klausimų kategoriją reikia pradėti pačiu svarbiausiu klausimu, kuriuo, prieš norint parengti kursą užsienio kalbai mokytis, siekiama sužinoti, kokios užsienio kalbos labiausiai domina studentus. Taigi į klausimą, kokią kalbą norėtumėte išmokti artimiausioje ateityje buvo atsakyta stebėtinai įdomiai, nes net 12 (30%) apklaustųjų norėtų išmokti anglų kalbos, galbūt galima manyti, kad nebūtinai išmokti pradėdant abėcėle, o tiesiog pagilinti savo žinias, kalbos įgūdžius, nes visiems gerai žinoma, jog mokykloje ir dažnose universitetų studijų programose privaloma mokytis anglų kalbą. Antrąją pagal poreikį išmokti norima – apklaustieji nurodė rusų kalbą, kurią norėtų mokytis 10 (25%) studentų, trečioje vietoje liko vokiečių kalba, kurią rinkosi 9 (22,5%) apklaustųjų. Mažiau svarbiomis kalbomis studentams pasirodė ispanų kalba (15% - 6 stud.) bei asmeniškai pasirinkti variantai, t.y. kinų kalba, vokiečių ir ispanų (kartu) ir prancūzų (7,5 % – 3 stud.) kalbos.

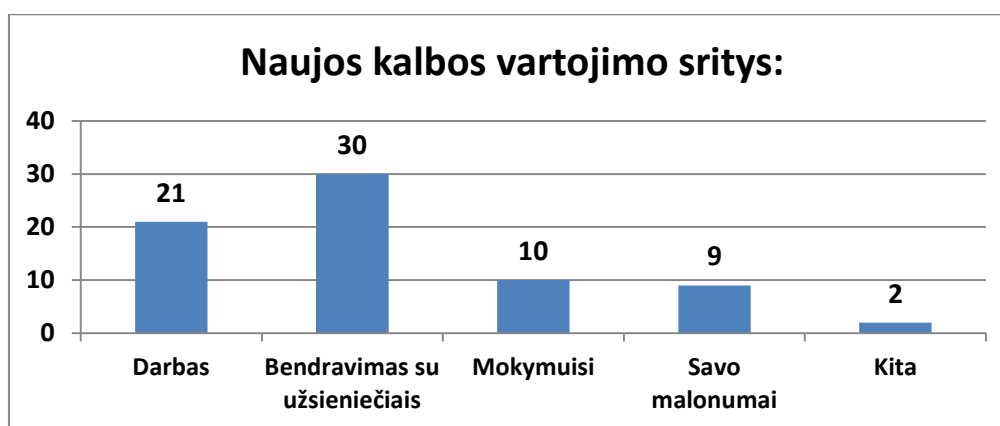
Svarbiu laikytas klausimas, kuriuo siekiama sužinoti, koks kalbų mokymosi metodas studentams atrodo efektyviausias. Pagal pateiktą diagramą matoma, jog studentų nuomonė šiuo klausimu buvo skirtinga.



**1 pav.** Efektyviausi kalbų mokymosi metodai

Iš atsakymų paaiškėjo, jog apklausoje dalyvavusiems studentams labiausiai priimtinais kalbos mokymosi metodais išlieka visiems gerai žinomas tradicinis metodas, kai mokytojas perduoda studentams informaciją tiesioginių susitikimų metu, moko ją pritaikyti (14 stud.) Mažiau patikimu laikytinas nuotolinio mokymosi metodas, kai mokymasis paremtas savarankišku darbu, mokymuisi naudojant elektroninę mokymosi medžiagą, skaidres, internetą ir kt. Likusieji 14 studentų (kaip ir pirmuoju atveju) pasisako už tai, jog tiek tradicinis, tiek nuotolinis mokymas gali privesti prie vienodų rezultatų, nes abu yra laikomi lygiaverčiais.

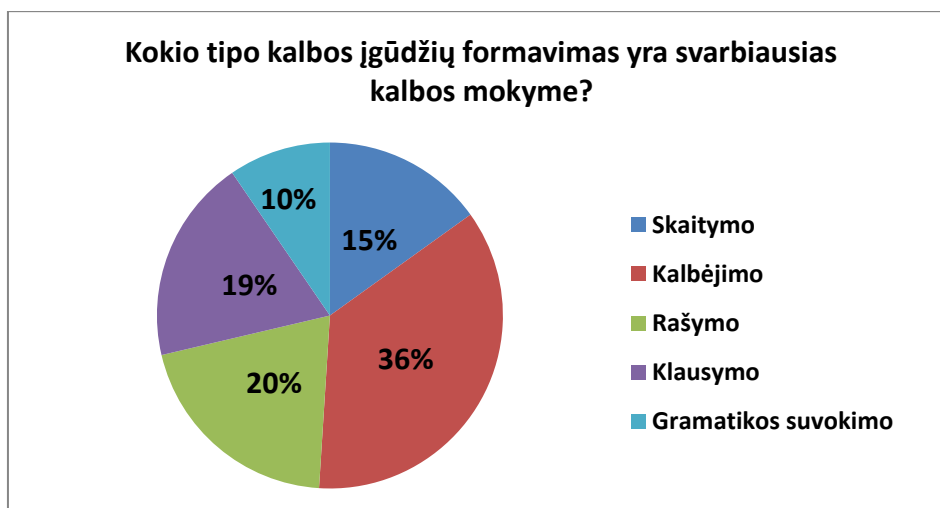
Sekančiu klausimu buvo norima sužinoti, kur naujai išmoktą užsienio kalbą studentai ketina vartoti.



**2 pav.** Kalbos vartojimo sritys

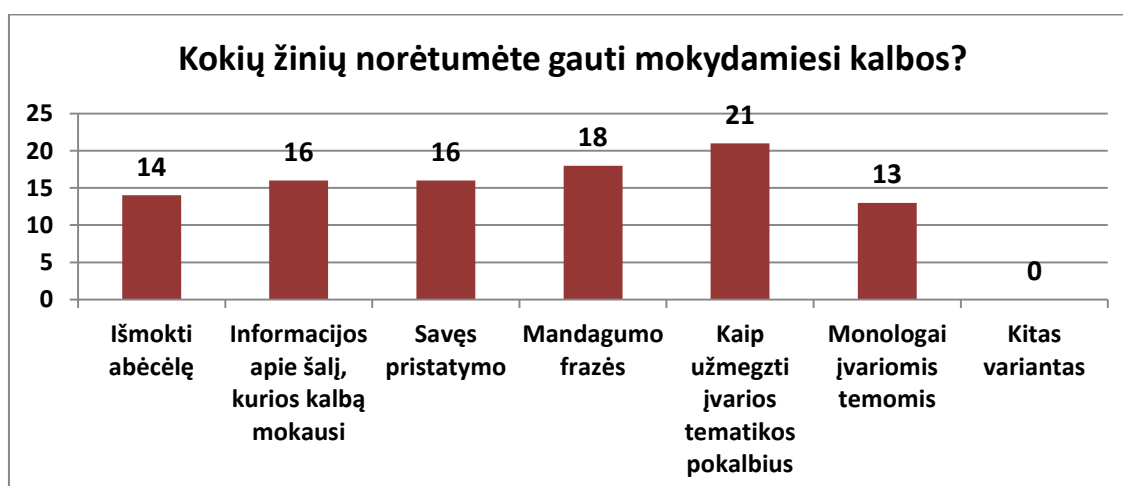
Remiantis gautais atsakymais paaiškėjo, jog studentams naujai išmokta užsienio kalba labiausiai reikalinga norint bendrauti su užsieniečiais, galima manyti, jog taip plečiamas akiratis. Taip pat vienu iš svarbiausių užsienio kalbos vartojimų sričių išlieka darbas, nes net 21 asmeniui prireikia svetimų kalbos žinių darbo aplinkoje. Likusieji nurodo, jog naujai išmoktą kalbą vartos mokymosi procese, savo malonumui ir pan.

Į klausimą, kokio tipo kalbos įgūdžių formavimas studentams yra svarbiausias (kokius įgūdžius norėtų suformuoti) dauguma atsakė, jog svarbiausiais išlieka kalbėjimo ir rašymo įgūdžiai, o mažiausiais svarbius įvardino gramatikos suvokimo. Skirtingas nuomones įrodanti skritulinė diagrama pateikiama žemiau.



**3 pav.** Kalbos įgūdžių formavimas

Taip pat, kad būtų galima orientuotis, kokių dalykų mokyti studentus, jeigu tikrai būtų nuspręsta kurti pasirinktos užsienio kalbos mokymo kursą, studentai buvo paklausti, kokių žinių būtų norima įgyti mokantis kalbos.

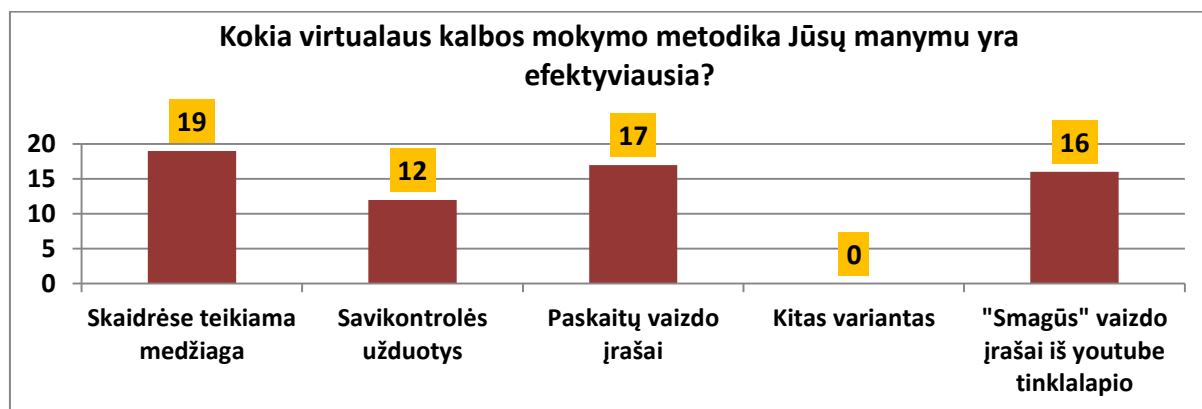


**4 pav.** Žinios mokymosi procese

Taigi dauguma (21 studentas) pasisakė, jog aktualiausia būtų išmokti užmegzti pokalbį su kitu žmogumi (buitinei kalbai būdingas bruožas, tad būtų logiška, kad studentas pradeda mąstyti nuo paprastų dalykų išmokimo). Taip pat ne mažiau svarbu mokėti elementaraus etiketo frazių, gebėti pristatyti save svetimiems žmonėms ir išmanyti svietimos kultūros peripetijas (informacija apie šalį). Mažiau svarbu studentams pasirodė mokytis abėcėlės (nebent tai būtų rusų kalbos abėcėlė) ar pasisakyti įvairiomis temomis/klausimais (monologai).

Pasirinktus dalykus, studentai geriausiai išmoktų skaitant tekstus (27,1% - 23 stud.), versdami tekstus (24,7% - 21 stud.). Nežymiai, bet mažiau naudos teiktų kalbėjimasis su kitu žmogumi (22,4% - stud.) ir dialogų rašymas (20% - 17 stud.). Kalbos mokymo metodika, kai mokymo procese taikomas įvairių užduočių, pateikiamų su atsakymų variantais atlikimas, studentams pasirodė nereikšmingas, todėl sulaukė tik 5 apklaustųjų dėmesio, kas sudaro 5,9%.

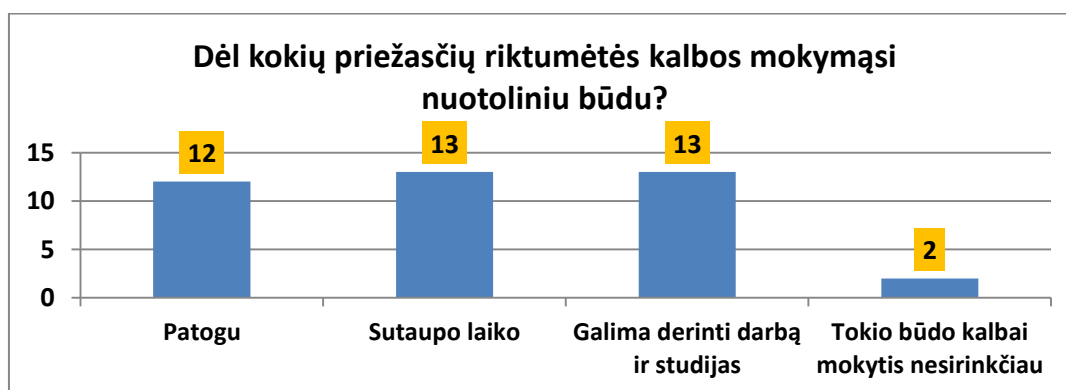
Paskutiniu metu su kalbos mokymu susijusiu klausimu norėta sužinoti, kokia virtualaus kalbos mokymo metodika studentams atrodo efektyviausia.



**5 pav.** Efektyviausia virtuali mokymosi metodika

Pastebėta, jog studentai prioritetą teikia skaidrėms, be kurių šiais laikais mokymo(si) procesas nevyksta. Naudojantis skaidrėmis dėstytojai gali lengviau pateikti informaciją studentams. Kadangi skaidrėse būna pateikta tik svarbiausia medžiaga, tik ta, kurią reikia žinoti, todėl studentui skaidrės tarsi išsigelbėjimas nuo papildomų informacijos šaltinių ieškojimo. Paskaitų vaizdo įrašais paremta metodika taip pat geras pasirinkimas, nes įrašą galima peržiūrėti daug kartų. Įrašuose vaizdas persipina su garsu. Kai vienu metu stimuliuojamos kelios žmogaus jausmų – paprasčiau įsiminti norimą informaciją. Suprantama, dažnai studentas įsimena tik „smagius“ dalykus, todėl įtraukti „smagios“ metodikos elementus į mokymą tikslinga. Mažiausiai naudos mokymosi (pasak diagramoje pasiskirsčiusių studentų nuomonių) teikia savikontrolės užduotys.

Galiausiai bus aptarti rezultatai, kuriuos galima priskirti prie NM. Apklausoje buvo klausama, dėl kokių priežasčių būtų pasirinkta mokytis užsienio kalbos nuotoliniu būdu.



**5 pav.** Priežastys NM pasirinkimui

Analizuojant diagramą matoma, jog vienodas skaičius (po 13) studentų rinkęsi netradicines studijas todėl, kad būtų galima jas suderinti su darbu ir tai sutaupyti laiko, nes, kaip jau buvo minėta anksčiau, mokymosi aplinkos ir laiko pasirinkimas – tai nuotoliniam mokymuisi būdingi bruožai, o mokantis tradiciniu būdu tektų lankyti kalbos kursus nustatytu laiku ir nustatytoje vietoje. NM patogumas yra 3-ioje vietoje, o 2 iš apklaustųjų apskritai tokio mokymosi būdo nesirinktų.



Visada svarbu pagalvoti ir apie tai, kur reikės talpinti kursą/medžiagą nuotoliniam mokymuisi, tad klausimas, kokioje virtualioje erdvėje norėtumėte gauti mokymosi medžiagą/informaciją į apklausą taip pat buvo įtrauktas. Kaip patogiausią tam vietą studentai išskiria bet kokią internetinę svetainę (44,4% - 20 stud.), kurioje būtų galima patalpinti medžiagą. Bet kokia čia būtų galima laikyti atsitiktinai parinktą, kursui paskirtą svetainę, į kurią būtų galima patalpinti medžiagą. 18 studentų (40%) pasisako už „Moodle“ sistemą. Tikėtina, kad dėl to, jog yra tekę su ja susidurti universitete (ieškant modulių, medžiagos ir kt.). Nedidelė dalis, t.y. 7 studentai (15,6%) neprieštarautų, jeigu mokymosi medžiaga būtų pateikiama iCloud (virtualaus darbalaukio) sistemoje. Galiausiai studentų buvo paklausta, ar kalbos mokymas nuotoliniu būdu būtų efektyvus. Pusė iš apklaustųjų (50% - 20 stud.) atsakė teigiamai, 3 (7,5%) neigiamai, kad bus iš dalies (šiek tiek) efektyvus mokymas galvoja 11 studentų (27,5%) ir net 15% pasisakė nežinantys, ar tai būtų efektyvu.

Taigi atlikta apklausa parodė, jog skirtingų universitetų studentai yra suinteresuoti išmokti daug naujų kalbų. Studentams tiek tradicinis, tiek nuotolinis mokymosi būdai atrodo vienodai priimtini. Naujai išmokta kalba, pasak studentų, būtų labiausiai pritaikoma bendraujant su užsienio piliečiais. Apklaustųjų nuomone didžiausią dėmesį mokantis reikia skirti kalbėjimo įgūdžiams lavinti, ką pavyktų padaryti, jeigu nuotolinio mokymo metu būtų mokoma, kaip užmegzti pokalbį įvairiomis temomis. Geriausi rezultatai, studentų nuomone, būtų pasiekti naudojantis skaidrėse pateikiama medžiaga, o patogiausia gauti medžiagą būtų iš tinklalapio (pritaikyto mokymuisi).

## **16 PRIEDAS DĖSTYTOJŲ APKLAUSOS KLAUSIMAI**

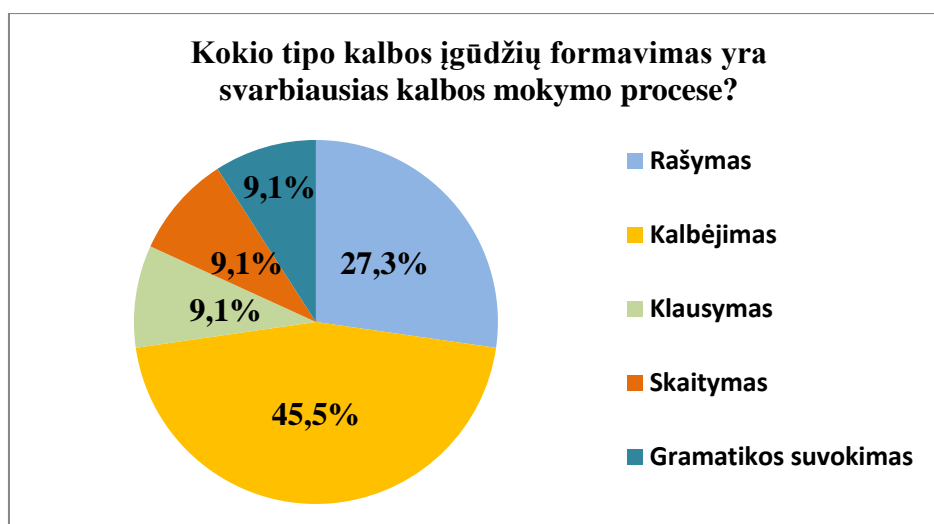
1. Kokio tipo kalbos įgūdžių formavimas yra svarbiausias kalbos mokymo procese?
2. Kokios priemonės kalbų mokymosi medžiagai teikia patrauklumo?
3. Kokia virtualaus kalbos mokymo metodika yra efektyviausia?
4. Ar tikslinga prieš mokant kalbos suteikti informacijos apie šalį?
5. Kokios temos yra neišvengiamos kalbos mokyme?
6. Kokio tipo kalbos užduotys yra efektyviausios?
7. Kokioje virtualioje erdveje patogiausia pateikti mokymosi informaciją?
8. Pastabos ir komentarai, į kuriuos reikia atsižvelgti kuriant kurso medžiagą / parenkant mokymo metodus?

## 17 PRIEDAS DĚSTYTOJŲ APKLAUSOS ANALIZĖ (PILNA)

Nepaisant to, jog studentai iš patirties gali įvardinti, kokios priemonės jiems gali būti naudingos mokymosi procese, verta atkreipti dėmesį ir į dėstytojų nuomonę, ką šie gali pasakyti ta pačia tema. Dėl šios priežasties buvo sugeneruota apklausa [apklausa.lt](http://apklausa.lt) tinklalapyje, kuria buvo apklausti 5 KTU lietuvių kalbos specialistai siekiant išsiaiškinti, kokius, pasak kalbos specialistų, kurie nuolatos susiduria su kalbos mokymu, jame taikomomis priemonėmis, metodus yra tikslingiausia taikyti mokant užsienio studentus lietuvių kalbos virtualiu būdu. Toks dėstytojų skaičius pasirinktas todėl, kad tai galima laikyti optimaliu kiekiu. Apklauskos anketą sudaro 8 klausimai, kurie ne detaliam apsprendžia svarbiausius punktus, į kuriuos reikia orientuotis prieš pradėdant kurti virtualų kalbos mokymosi kursą. Apklauskos platinimas vyko naudojantis bakalauro mokymosi metais sukauptais dėstytojų kontaktais.

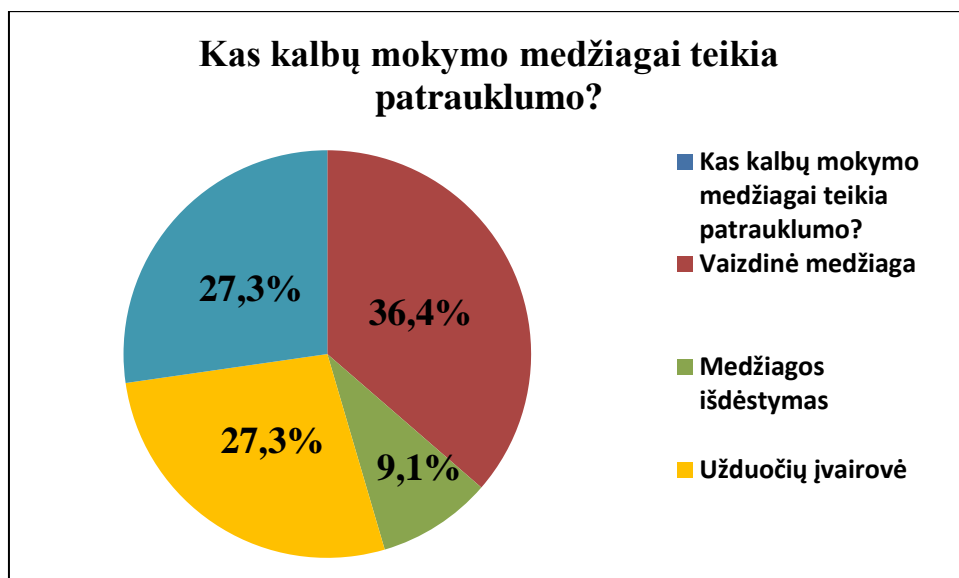
Tiek studentų, tiek dėstytojų apklauskos tikslas yra vienodas, t.y. sužinoti, kokie kalbos mokymosi principai būtų efektyviausi teikiant virtualų kursą, tačiau dėstytojams skirtoje apklausoje nepateikiami bendrojo pobūdžio klausimai, klausimai apie kalbos vartojimo patogumą bei sritis, ir kita. Apklausoje vyrauja klausimai, tiesiogiai susiję su kalbos mokymo procesu bei reikalingomis priemonėmis ir kitomis minėtojo proceso detalėmis.

Visų pirma, dėstytojams skirtoje apklausoje klausiama, kokio tipo kalbos įgūdžių formavimas yra svarbiausias kalbos mokymo procese. Svarbiausiais įgūdžiais yra laikomi kalbėjimo bei rašymo įgūdžiai. Tokios pačios nuomonės laikosi ir studentai. Mažiausiai svarbiais yra laikomi visi likusieji įgūdžiai, t.y. klausymo, skaitymo bei gramatikos suvokimo. Skirtingą nuomonių pasiskirstymą įrodanti skritulinė diagrama pateikiama žemiau.



6 pav. Įgūdžiai kalbos mokymosi procese

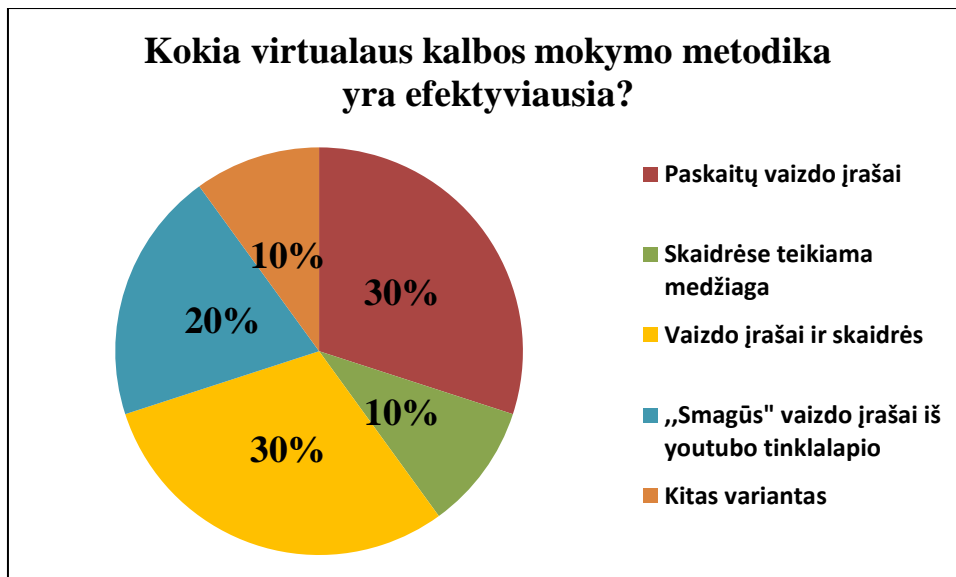
Antruoju anketos klausimu norėta sužinoti, kokios priemonės kalbų mokymosi medžiagai teikia patrauklumo.



**7 pav.** Kalbos mokymosi medžiagos pateikimas

Analizuojant skritulinę diagramą matoma, jog vaizdinės medžiagos teikimas mokymosi procese yra laikomas svarbiausiu aspektu, kuris šiuo atveju žymi 36,4% diagramos dalies (pažymėjo 4 iš 5 dėstytojų). Vienodai patraukliu teikimo būdu (27,3%) yra laikoma teikiamų užduočių įvairovė bei teikimo paprastumas.

Dar vienu apklausos klausimu siekta sužinoti, kokia virtualaus kalbos mokymo metodika dėstytojams atrodo efektyviausia.

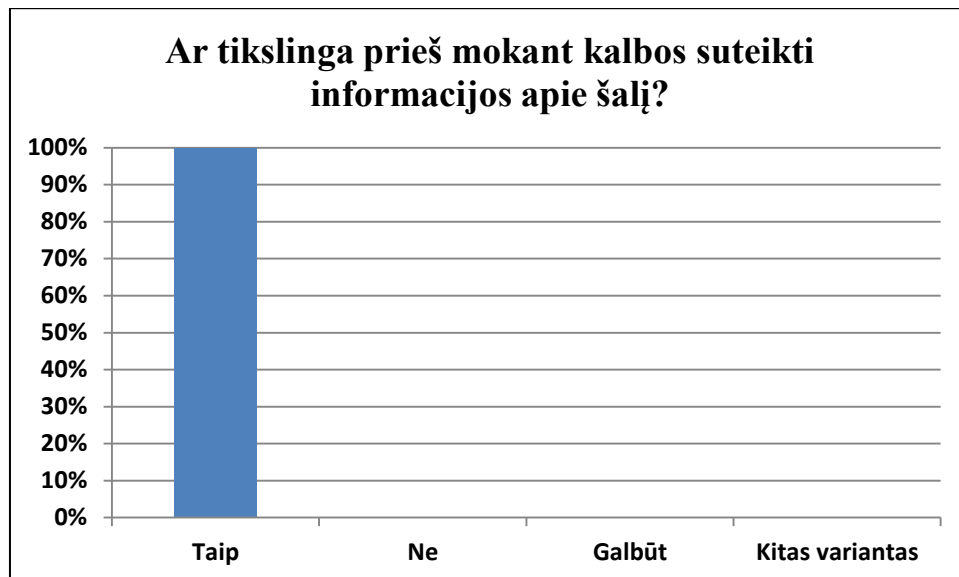


**8 pav.** Efektyviausios kalbos mokymosi metodikos

Pastebėta, kad dėstytojai yra linkę manyti, jog efektyviausia mokymo metodika remiasi paskaitų vaizdo įrašų ir skaidrių teikimu. Studentai šiuo klausimu prioritetą teikia skaidrėms, kitas priimtinas atvejis – paskaitų vaizdo įrašais paremta metodika. Antroje pozicijoje dėstytojai palieka „smagius“ vaizdo įrašus, kuriems aukštą poziciją taip pat priskiria ir studentai, nes, kaip buvo minima anksčiau, dažnai studentas įsimena tik „smagius“ dalykus, todėl įtraukti „smagios“ metodikos elementus į mokymą tikslinga. Mažiausiai naudos mokymosi procese gali teikti tik

skaidrių teikimu paremta mokymo metodika, bei variantu „kita“ įrašytas atsakymas „Youtube vaizdeliai“. Mažiausiai naudos mokymesi, pasak diagramoje pasiskirsčiusių nuomonių, teikia savikontrolės užduotys.

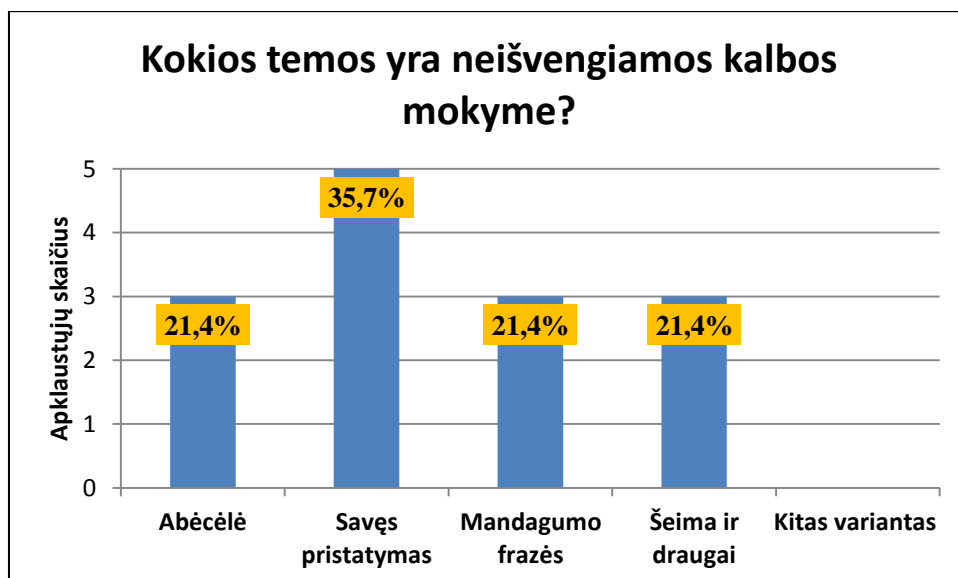
Ketvirtuoju anketos klausimu norėta sužinoti, kokia yra dėstytojų nuomonė apie tai, kad teikiant kursą būtų teikiama informacija apie šalį, kurios kalbos bus mokomasi.



**9 pav.** Informacijos apie šalį teikimas

Neabejotinai į šį klausimą visi apklaustieji atsakė vienareikšmiškai – taip, tikslinga. Informaciją apie šalį yra tikslinga pateikti, nes studijuojantieji privalo žinoti bent pačius svarbiausius faktus ir galbūt keletą pagrindinių savybių, kurios būdingos šaliai.

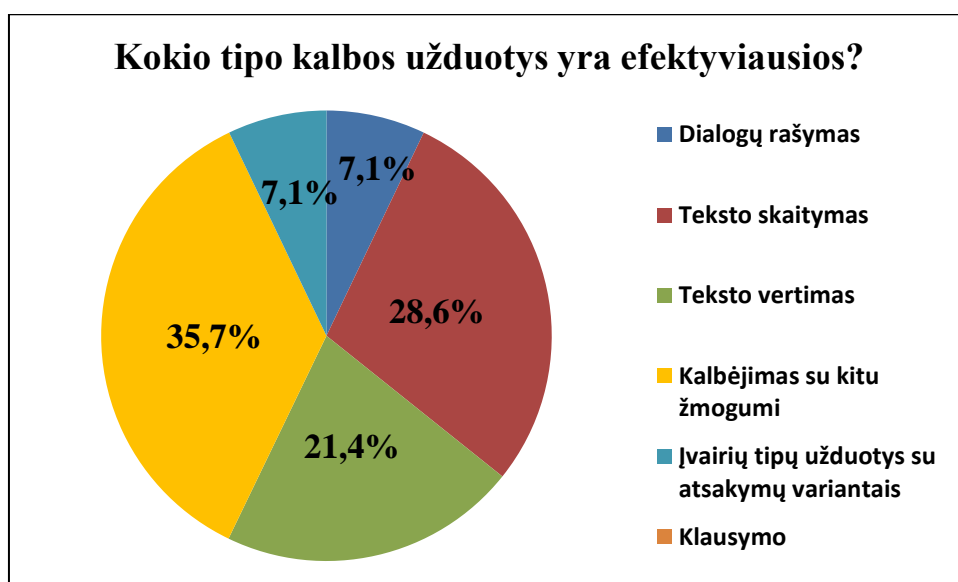
Taip pat, kad būtų galima orientuotis, kuria kryptimi vykdyti mokymus, kokios informacijos teikimas kurse būtų tikslingas, buvo klausama, kokios temos yra neišvengiamos kalbos mokyme. Studentai į šį klausimą atsakė, jog aktualiausia būtų išmokti užmegzti pokalbį su kitu žmogumi. Ne mažiau svarbu mokėti elementaraus etiketo frazių, gebėti pristatyti save svetimiems žmonėms ir išmanyti svetimą kultūrą (informacija apie šalį). Mažiausiai svarbu studentams pasirodė mokytis abėcėlės ar išreikšti nuomonę įvairiomis temomis/klausimais (monologai).



**10 pav.** Tematika kalbos mokymosi procese

Tačiau dėstytojai šiuo klausimu išreiškė kitokią nuomonę nei studentai. Svarbiausiu mokymo aspektu jie išskiria savęs pristatymo, o likusias temas apie abėcėlę, šeimą ir mandagumą paliko antroje vietoje.

Pasirinktus dalykus, pasak dėstytojų, geriausiai studentai įsisavintų jeigu kalbėtūsi su kitais žmonėmis svetima kalba (~38% – 5 dėst.), skaitytų tekstus (~29% – 4 dėst.) bei verstų tekstus (~21% – 3 dėst.). Mažiausiai reikšmingi yra įvairių dialogų rašymai ir užduotys, kurios turi variantus (pagalbiniai variantai – lengviau atsakyti), kas sudaro po ~7% (po 1 dėst.).

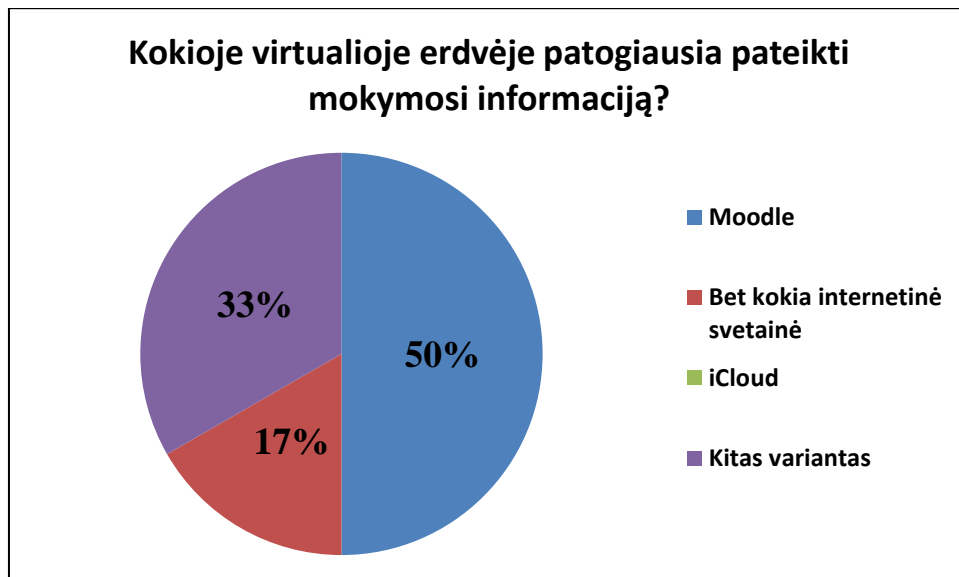


**11 pav.** Efektyvių kalbos užduočių tipai

Lyginat dėstytojų nuomonę su studentų galima teigti, kad lūkesčiai išsiskiria, nes studentai linkę manyti, jog pasirinktus dalykus jie geriausiai išmoktų skaitant tekstus (27,1% – 23 stud.), versdami tekstus (24,7% – 21 stud.). Kalbos mokymo metodika, kai mokymo procese taikomas

įvairių užduočių, pateikiamų su atsakymų variantais atlikimas, studentams pasirodė nereikšmingas, todėl sulaukė tik 5 apklaustųjų dėmesio, kas sudaro 5,9%.

Kuriant virtualų kursą reikia pagalvoti ir apie tai, kokioje aplinkoje būtų naudingiausia jį realizuoti. Taigi paskutinis reikšmingas klausimas dėstytojams, kokioje virtualioje erdvėje patogiausia teikti mokymosi informaciją.



**12 pav.** Patogiausia VM erdvė

Vertinant nuomones pagal diagramą, didžiausią dalį (50%) užima virtuali mokymosi aplinka „Moodle“. Antroje vietoje, klausime su pasirenkamuoju atsakymo variantu, dėstytojų pasirinkimu yra „Adobe connect“ aplinka arba kita (neįvardinta kokia) mokymosi aplinka. Tai galėtų būti „first class“ ir pan. Mažiausią dalį (33% – 1 dėst.) sudaro nuomonė, jog informacija gali būti talpinama tiesiog pasirinktoje internetinėje svetainėje.

## 18 PRIEDAS UŽSIENIO STUDENTŲ APKLAUSOS KLAUSIMAI

1. Lytis.
2. Amžius.
3. Šalis, iš kurios atvykote studijuoti.
4. Mokymosi įstaiga.
5. Esate dirbantis / nedirbantis studentas?
6. Kokią kalbą norėtumėte išmokti ateityje?
7. Kokie yra efektyviausi kalbos mokymosi metodai?
8. Naujos kalbos vartojimo sritys.
9. Kas mokymosi medžiagai teikia patrauklumo?
10. Kokių įgūdžių pageidaujate įgyti mokydami lietuvių kalbos?
11. Kokioms kalbos mokymosi užduotims skirate pirmenybę?
12. Kokios kalbos mokymosi užduotys yra nereikalingos mokantis kalbos?
13. Kokie yra efektyviausi kalbos mokymosi metodai?
14. Patogiausia virtuali mokymosi terpė mokymosi medžiagai talpinti.
15. Kodėl anksčiau neišbandėte virtualaus mokymosi?
16. Dėl kokių priežasčių rinktumėtės virtualų mokymąsi?

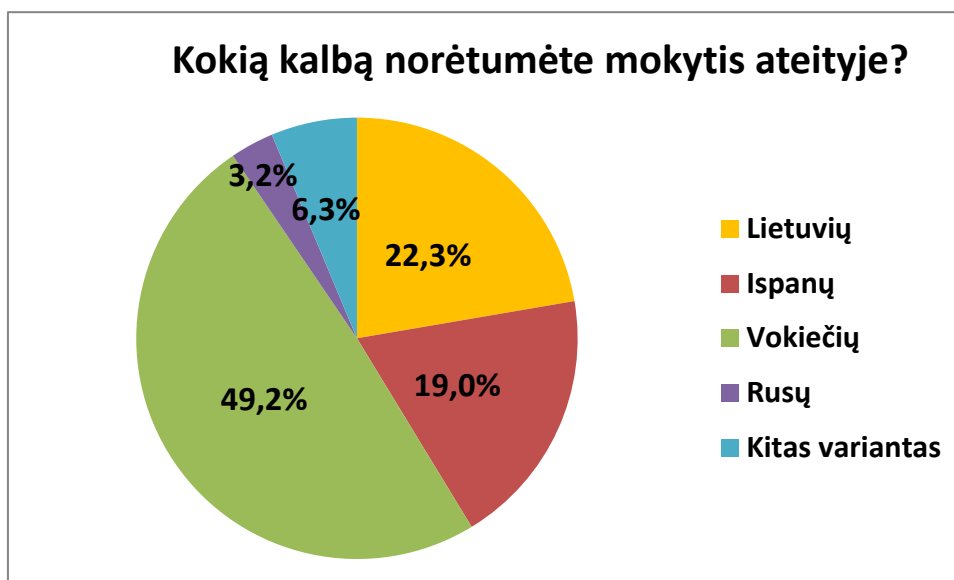


## 19 PRIEDAS UŽSIENIO STUDENTŲ POREIKIŲ ANALIZĖ (PILNA)

Siekiant ištirti užsienio studentų poreikį studijuoti svetimą užsienio kalbą nuotoliniu būdu, specialiai užsienio studentams buvo sugeneruota apklausa anglų kalba [apklausa.lt](http://apklausa.lt) tinklalapyje. Apklausa buvo pateikta praėjusių metų rudens semestre studentams, studijuojantiems Kauno technologijos universitete. Apklausa išplatinta bendradarbiaujant su viena iš KTU universiteto dėstytojų, kuri dėsto paskaitas užsienio studentams. Vienos iš dėstytojos paskaitų metu studentams buvo pateikta užduotis – atlikti apklausą. Tokiu būdu išplatinus apklausą pavyko sulaukti grįžtamojo ryšio iš 35 užsienio studentų. Visi atsakiusieji į apklausos klausimus yra pirmo kurso studentai, 5 yra Erasmus programoje dalyvaujantys asmenys, o likusieji – nuolatinių KTU universiteto studijų studentai.

Apklausos anketą sudaro 16 klausimų. Klausimus galima suskirstyti į bendrojo pobūdžio (lytis, amžius, mokymosi įstaiga ir pan.), susijusius su mokymosi medžiaga ir jai teiktiniais pageidavimais, taip pat yra du klausimai apie nuotolinį mokymąsi. Pirmiausia bus aptariami bendrojo pobūdžio klausimai. Kaip jau buvo minima anksčiau, atsakymus į apklausos klausimus pateikė 35 užsienio studentai. Šiuo atveju tai yra 29 (82,9%) vyrai ir tik 6 (17,1%) moterys. Visi apklaustieji pažymėjo esantys asmenys iki 25 metų amžiaus. Toliau apklaustųjų buvo prašoma nurodyti, iš kokių šalių jie yra atvykę studijuoti į Lietuvą. Didžioji dalis, net 25 (71,4%) studentai, teigia esantys iš Indijos, 3 (8,6%) iš Ispanijos, o likusieji 7 (20%) savo gimtąją šalimi nurodė esant Italiją (2 stud.), Braziliją (1 stud.), Latviją (1 stud.), Azerbaidžaną (1 stud.) ir Libaną (2 stud.). Tam, kad užtikrintai būtų galima teigti, jog visi studentai, dalyvaujantys apklausoje tikrai yra KTU bendruomenės nariai, o ne laisvieji klausytojai, buvo paprašyta nurodyti, kokiame universitete jie studijuoja. Visi apklaustieji 100% nurodė esantys KTU studentai. Iš jų net 33 (94,3%) yra nurodę esantys tik studentai (niekur nedirba), o likusieji į klausimą atsakymo nepateikė.

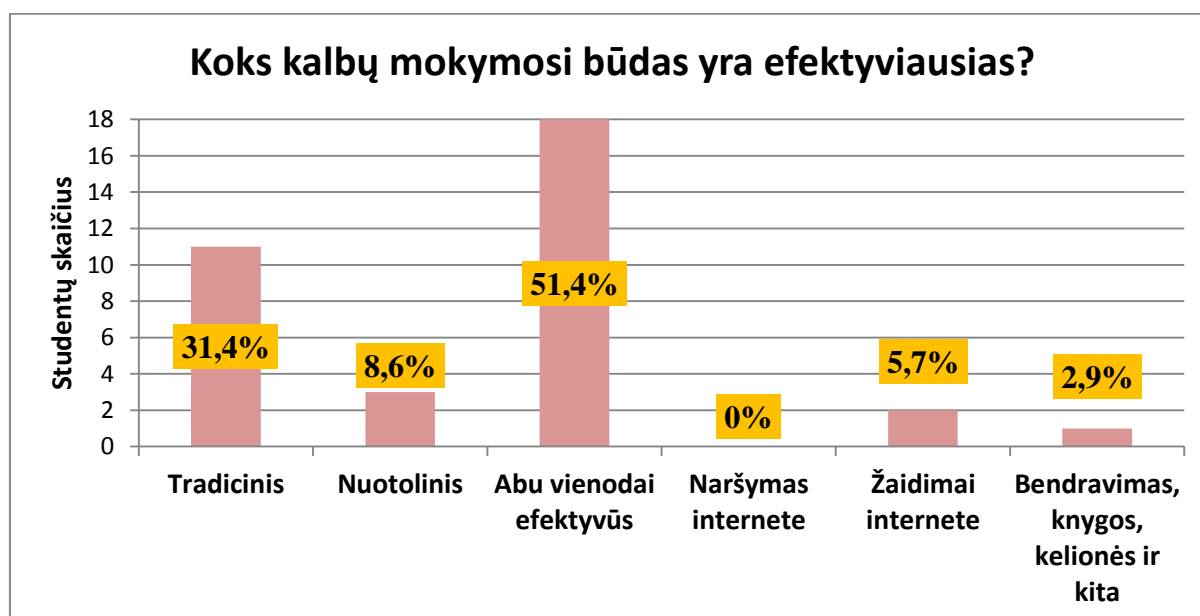
Su kalba susijusių klausimų kategorija pradedama klausimu studentams, kokią kalbą norėtumėte išmokti artimiausioje ateityje. Šiame klausime buvo galima pažymėti kelis pageidaujamus atsakymo variantus, kas suformavo tokius atsakymo rezultatus:



**13 pav.** Pageidautina išmolti kalba

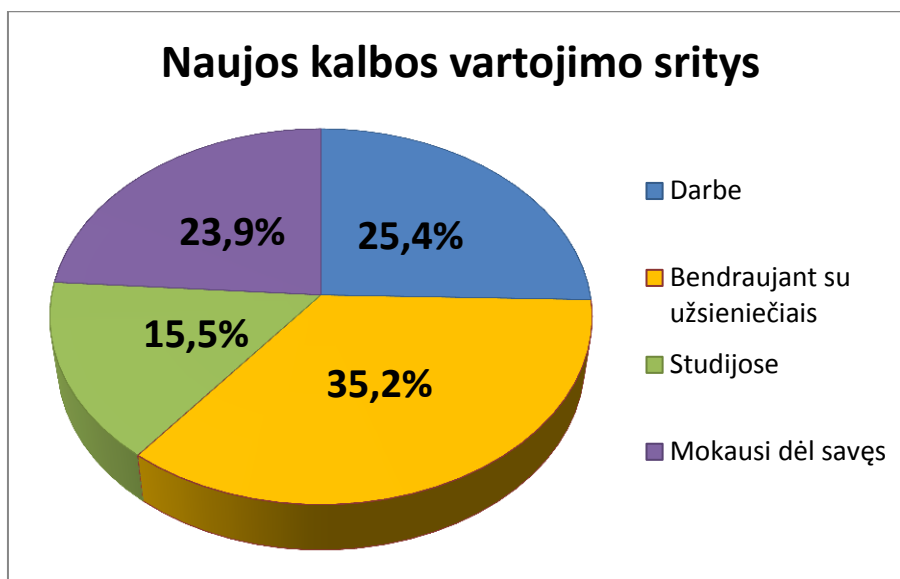
Laikiniai (galbūt ir nuolatos) gyvendami Lietuvoje 31 stud. (49,2%) nurodė pageidaujantys išmolti ne lietuvių, bet vokiečių kalbos ypatumų. Tai nebūtų keistas reiškinys, jeigu besimokantieji būtų 3 arba 4 kurso studentai, kas reikštų, jog per kelis metus, besimokant Lietuvoje studentai jau yra įgiję lietuvių kalbos žinių, todėl dabar pageidauja išmolti kitų užsienio kalbų. Tačiau žinant faktą, jog apklaustieji yra pirmakursiai ir teikia pirmenybę ne lietuvių, šalies, kurioje studijuojama, o kitai užsienio kalbai, toks faktas stebina. Lietuvių kalbos žinių įgijimas domina 14 (22,3%) apklaustųjų, 12 stud. (19%) domina ispanų kalba, 2 stud. (3,2%) – rusų kalba, o kiti 4 stud. (6,35%) įvardijo tokias kalbas kaip prancūzų, arabų, italų ir portugalų.

Svarbiu klausimu, kaip ir lietuvių studentų poreikių apklausoje, yra laikomas klausimas apie efektyviausią kalbos mokymosi būdą. Pagal pateiktą diagramą matoma, kaip pasiskirstė studentų nuomonės.



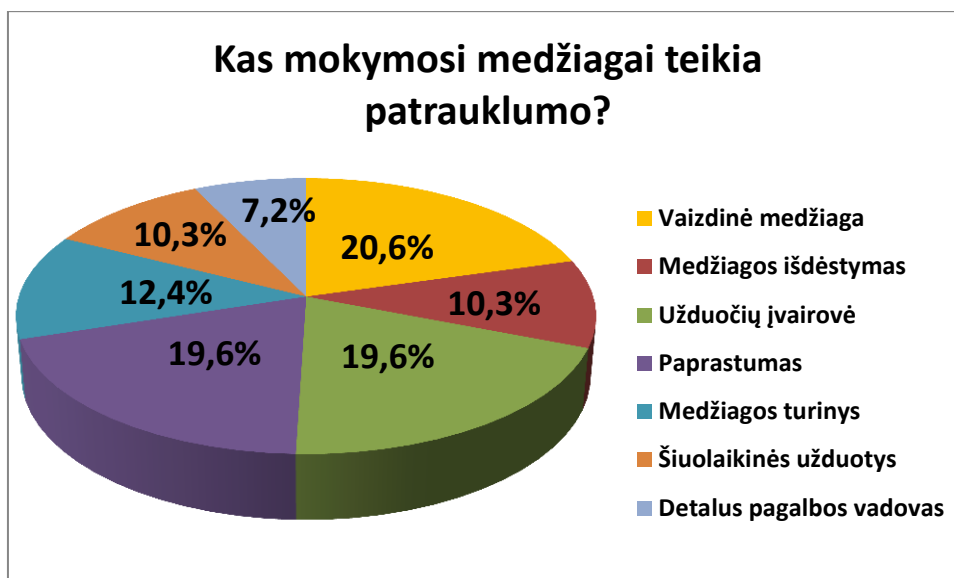
**14 pav.** Efektyviausias kalbos mokymosi būdas

Iš studentų atsakymų paaiškėjo, kad tiek tradicinis, tiek nuotolinis mokymasis yra lygiaverčiai ir vienodai efektyvūs besimokantiejiems – taip teigia 18 iš 35 (51,4%) studentų. Apklaustųjų nuomone, tradicinis mokymasis užima tik antrą vietą, šį variantą pasirinko 11 iš 35 (31,4%) studentų. Nuotolinį mokymąsi kaip efektyvų renka 3 (8,6%) studentai. Iš asmeninės praktikos žinoma, kad dažnai kalbos mokymuisi pasirenkami netradiciniai metodai, tokie kaip kalbos lavinimas naršant internete, žaidžiant kompiuterinius žaidimus ir pan. Šiuo atveju 2 stud. (5,7%) sėkmingu kalbos įgūdžių įgijimo metodu laiko įvairius žaidimus internete, 1 stud. (2,9%) – bendravimą su kitakalbiais, knygų skaitymą ir keliavimą, o variantą „naršymas internete“ (tai gali būti straipsnių skaitymas, bendravimas ir pan.) efektyviu metodu nelaiko nė vienas studentas. Kitu klausimu buvo norima sužinoti, kur ir kokiose situacijose naujai išminktą užsienio kalbą studentai ketina vartoti.



**15 pav.** Kalbos vartojimo sritys

Iš atsakymų paaiškėjo, kad dauguma studentų naujai išminktą kalbą žada vartoti bendraudami su užsieniečiais (25 stud. – 35,2%), darbe (18 stud. – 25,4%), daugiau kaip pusė įvardija, kad kalbos mokėjimas jiems reikalingas tiesiog akiračiui praplėsti (17 stud. – 23,9%). Likusiems kalbos mokėjimas yra reikalingas sėkmingoms studijoms (11 stud. – 15,5%). Galima daryti prielaidą, kad tie studentai, kuriems reikalinga pasirinkta kalba studijose, nori stabiliai jaustis ir pačioje Lietuvoje. Dar vienu klausimu norėta sužinoti, kas, užsienio studentų nuomone, kalbų mokymo medžiagai gali teikti patrauklumo.



**16 pav.** Mokymosi medžiagos pateikimas

Iš surinktų atsakymų pastebėta, kad patrauklumo kalbų mokymo medžiagai, pasak studentų, gali suteikti pirmiausia iliustracinė medžiaga (20 stud. – 20,6%), taip pat jos pateikimo ir naudojimo paprastumas (19 stud. – 19,6) bei užduočių įvairovė (19 stud. – 19,6%). Maždaug trečdalis apklaustųjų galvoja, jog mažiau svarbūs dalykai kalbos mokymo medžiagos patrauklumui yra medžiagos turinys (12 stud. – 12,4%), šiuolaikinės užduotys (10 stud. – 10,3%) ir medžiagos išdėstymas (10 stud. – 10,3%). Mažiausiai studentams svarbus mokymosi medžiagoje – pagalbos vadovas (7 stud. – 7,2%). Dar vienu klausimu norėta sužinoti, kokius svetimų kalbos įgūdžius studentai norėtų suformuoti mokydami. Dauguma studentų atsakė, jog svarbiausia yra mokėti užmegzti pokalbį įvairiomis temomis (20 stud. – 24,1%) bei prisistatyti (19 stud. – 22,9%). Antrai pozicijai galima priskirti tokius atsakymo variantus: mandagumo frazių išmokymas (14 stud. – 16,9%) ir informaciją apie kultūrinius šalies aspektus (11 stud. – 13,3%). Mažiausiai svarbūs kalbiniai aspektai, pasak studentų, yra abėcėlės mokymasis (9 stud. – 10,8%), įvairių temų leksikos praplėtimas (9 stud. – 10,8%) ir bendravimas su kitais žmonėmis (1 stud. – 1,2%).

Toliau apklausoje pateikiama klausimų, kuriais studentai gali išreikšti savo nuomonę, į kurią būtų tikslinga atsižvelgti teikiant praktines užduotis nuotoliniame kalbos mokymosi kurse. Į klausimą, kokio tipo užduotims besimokantieji skiria pirmenybę, respondentų buvo atsakyta – kalbėjimo (29 stud. – 34,9%). Šiuo atsakymu galima patvirtinti visuomenėje vyraujančią nuomonę, jog svarbiausia yra mokėti susikalbėti su kitu asmeniu, kai skaitymo ir rašymo įgūdžiai yra nuvertinami. Antroje vietoje išlieka klausymo užduotys (18 stud. – 21,7%). Mažiausiai svarbiomis kalbos mokymosi užduotimis yra laikomas skaitymas (13 stud. – 15,7%), gramatikos išmokymas (12 stud. – 14,5%) ir rašymas (10 stud. – 12%). Tik vienas studentas pasirinko variantą, kuriame asmeniškai išreiškė savo nuomonę, kad visi klausimyne pateikti variantai yra vienodai svarbūs, t.y. visų išvardintų mokymosi veiklų derinimas gali duoti pačių geriausių kalbos išmokimo rezultatų. Kitu klausimu buvo paprašyta atsakyti, kokio tipo užduotys, užsienio studentų nuomone, yra

mažiausiai svarbios mokymosi procese. Paradoksalu šioje situacijoje yra tai, kad pasirenkamasis variantas, kuriame studentai išreiškė nuomonę, jog viskas yra svarbu kalbos mokyme, tiek rašymas, tiek klausymas, tiek kiti kalbiniai dalykai (11 stud. – 22,9%) yra tokios pat svirties kaip ir rašybos išmokimas (11 stud. – 22,9%). Taip pat beveik trečdalis studentų (10 – 20,8%) galvoja, jog gramatika kalbos mokyme yra nereikalinga. Tokia nuomonė, tikėtina, gali būti susiformavusi dėl žemų rezultatų atliekant gramatines kalbos užduotis. Mažiausiai nereikalingomis besimokantieji įvardija klausymo (7 stud. – 14,6%), kalbėjimo (5 stud. – 10,4%) ir skaitymo (4 stud. – 8,3%) užduotis. Studentai pateikė savo nuomonę, kuri buvo svarbi siekiant sužinoti, kokia mokymosi medžiaga jiems priimtinausia mokymosi procese, kokiems pratimams jie skiria didžiausią dėmesį, todėl toliau seka klausimas apie kalbos mokymosi metodus, kurie iš jų yra efektyviausi. Nuomonės pasiskirstė skirtingai: efektyviausias metodas/technika norint išmokti kalbos yra dialogai tarpusavyje (23 stud. – 24,5%), edukaciniai žaidimai (19 stud. – 20,2%) ir vaizdo įrašų/filmų žiūrėjimas (18 stud. – 19,1%). Mažiausiai efektyvia mokymosi technika besimokantieji įvardina esant skaidres (13 stud. – 13,8%), vaizdo įrašų iš „Youtube“ peržiūra (11 stud. – 11,7%) ir savikontrolės testus (9 stud. – 9,6%). Į šį klausimą, kaip ir ankstesniu atveju, buvo sulaukta nuomonės iš vieno studento, kuris galvoja, jog reikia kombinuoti visas išvardintas užduotis, siekiant geriausio kalbos išmokimo rezultato.

Kaip ir dėstytojų apklausos atveju, užsienio studentų taip pat buvo klausiama apie patogiausią terpę kalbos medžiagai talpinti. Studentai atsakė lygiai taip pat kaip ir dėstytojais – patogiausia terpė yra „Moodle“, su tuo sutinka didžioji dalis studentų (30 stud. – 45,5%). Skirtingai nei dėstytojais, kurie antruoju atveju pasirinko bet kokią kitą mokymosi aplinką, studentai renkasi galimybę gauti medžiagą iš paties dėstytojo (26 stud. – 39,4%). Mažiausiai studentų susidomėjimo sulaukė bet kokio tinklalapio pasirinkimas, kuriame galima talpinti medžiagą (10 stud. – 15,2%). Šiuo atveju nuomonė sutampa su dėstytojų nuomone.

Paskutiniais dviem klausimais buvo pasidomėta studentų nuomone apie patį nuotolinį mokymąsi. Pirmuoju iš klausimų buvo paklausta, kodėl studentai niekada nėra bandę nuotolinių kalbos mokymosi kursų. Pagal vyraujančią nuomonę studentai galvoja, jog tokie kursai yra abejotinos kokybės (13 stud. – 36,1%), kiti tokio tipo kursų nesirenka dėl žinių apie nuotolinį mokymąsi bei kursų trūkumo (10 stud. – 27,8%), tretieji apie tokias studijas ir kursus apskritai nieko nėra girdėję (9 stud. – 25%), galiausiai 4 studentai (11,1%) nurodė, kad tokie kursai teikia mažai motyvacijos ir kalbos mokymesi būtinai reikalingas mokytojas. Paskutiniu klausimu buvo klausiama, dėl kokių priežasčių studentai ryžtųsi išmėginti nuotolines kalbos studijas/kursus. Didžioji apklaustųjų dalis nurodė, kad tokio kalbos mokymosi būdo nesirinktų (15 stud. – 42,9%), kitiems atrodo, kad tai puikus būdas suderinti kursus su darbu (11 stud. – 31,4%), tretiems tai atrodo

patogus mokymosi būdas (8 stud. – 22,9%), o likęs 1 studentas (2,9%) yra įsitikinęs, kad kalbos išmokimas yra nepaprastas procesas, todėl jį tai reikia žiūrėti atsakingai.

Atlikta apklausa parodė, jog į KTU atvykę studentai užsieniečiai yra suinteresuoti išmokti daug naujų kalbų, taip pat ir lietuvių kalbos. Priimtinausi kalbų mokymosi metodai studentams yra tradicinis ir nuotolinis, kurie yra vienodai efektyvūs. Naujai išmoktą kalbą užsienio studentai pageidauja vartoti bendraujant su užsieniečiais, taip pat jie norėtų suformuoti bendravimo įgūdžius. Studentų nuomone, svarbiausia kalbų mokymosi medžiagoje yra paprastumas, užduočių įvairovė ir vaizdinė medžiaga.

## 20 PRIEDAS KOKYBĖS TYRIMO APKLAUSOS KLAUSIMAI

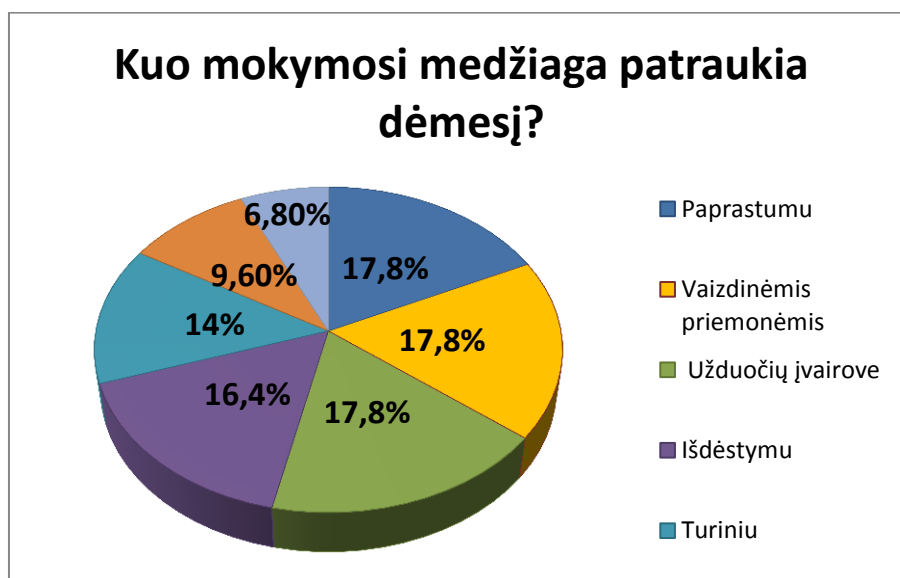
1. Lytis.
2. Amžius.
3. Šalis, iš kurios atvykote studijuoti.
4. Kuo mokymosi medžiaga patraukia dėmesį?
5. Ar priimtinas dvikalbystės naudojimas virtualiame kurse?
6. Ar Jums patiko mokymosi medžiagos išdėstymas kurse?
7. Ar naudingas mokymosi eiliškumo reguliavimas naudojant žingsnius?
8. Ar Jums patiko pateiktų užduočių įvairovė?
9. Ar naudinga kalbos mokytis naudojantis kalbos atpažinimo technologijomis?
10. Įvertinkite mokymosi medžiagą 1-100 balų sistemoje.
11. Kokie yra svarbiausi kurso privalumai?
12. Kokie yra kurso trūkumai?
13. Ar kursas yra įdomus ir išskirtinis?
14. Ar radote visų tipų užduotis, kurios reikalingos kalbai mokytis?
15. Ar norėtumėte išmokti kitų užsienio kalbų naudojantis tokio tipo kursu kaip tą, kurią vertinote ir išbandėte?

## 21 PRIEDAS KOKYBĖS TYRIMO ANALIZĖ (PILNA)

Prieš kuriant virtualų kursą buvo atliekamos studentų ir ekspertų apklausos, siekiant išsiaiškinti lūkesčius kuriamam kalbos mokymosi kursui. Priimtiems sprendimams dėl kalbos mokymosi kurso kūrimo įvertinti, kalbos mokymosi kursas pateiktas Kauno technologijos universiteto Informatikos fakulteto bei Socialinių, humanitarinių mokslų ir menų fakulteto užsienio studentams. Apklausa išplatinta bendradarbiaujant su minėtųjų fakultetų dėstytojais. Apklausa sudaryta iš 15 klausimų.

Atsakymus į apklausos klausimus pateikė 19 kursą išbandžiusių užsienio studentų, t.y. 11 (57,9%) vyrų ir 8 (42,1%) moterys. Didžioji dalis, net 10 (52,6%) studentų, teigia esantys iš Indijos, po 3 studentus (15,8%) iš Ispanijos ir Portugalijos, o likusieji 2 (10,5%) gimtąja šalimi nurodė esant Turkiją bei 1 (5,3%) Kiniją. Apklaustųjų amžius yra iki 25 metų.

Pirmiausia studentų buvo paklausta, kuo mokymosi medžiaga patraukia dėmesį. Rezultatai pasiskirstė taip: paprastumu, vaizdinėmis priemonėmis bei užduočių įvairove (po 17,8%), medžiagos išdėstymas (16,4%), medžiagos turinys (13,7%), užduočių šiuolaikiškumas (9,6%) ir detalus informacijos pateikimas (6,8%). Nuomonių pasiskirstymas pavaizduotas 17 pav.



17 pav. Mokymosi medžiagai patrauklumą teikiantys aspektai

Dvikalbystę visame kurse studentai vertina taip pat skirtingai. Daugumai, net 12 (60%) tai yra priimtina, 5 (25%) tai nėra priimtina, o likusiems 3 (15%) nesvarbu, ar medžiaga ir jos išdėstymas yra teikiamas viena, ar dviem kalbomis.

Kalbinės medžiagos išdėstymas kurse, taip pat ir mokymosi procesą reguliuojantys nurodomieji žingsniai yra vienodai priimtini visiems 19-kai studentų (po 100%).

Kurse vietoj gairių naudojami vadinamieji „žingsniai“. Į klausimą, ar naudingas žingsnių naudojimas prie kiekvienos temos, visi studentai atsakė teigiamai.

Kitu klausimu buvo norima sužinoti, ar studentams patiko pateiktų užduočių įvairovė. Į šį klausimą kaip ir į ankstesnįjį visi studentai atsakė teigiamai.



Toliau buvo paklausta apie kalbinių įgūdžių tobulinimą naudojant kalbos atpažinimo technologijas. Visi apklaustieji vienareikšmiškai sutiko, jog tai yra naudingas dalykas.

Paprašius studentų įvertinti pateiktą mokymosi medžiagą vertinimo skalėje nuo 1 iki 100, balai pasiskirstė įvairiai: 10 studentų mokymosi medžiagą vertina 90-100 balais, 6 galvoja, jog mokymosi medžiaga verta 80-90 balų, ir likusieji trys studentai vertina 70-80, 60-70 ir 20-30 balų sistemoje.

Taip pat studentų buvo paprašyta įvardinti labiausiai naudą teikiančius kurso dalykus. Daugiausiai, net 11, studentų teigia, kad viskas kurse jiems yra naudinga, 2-iems studentams – kalbinius įgūdžius lavinančios užduotys, o likusiems: medžiagos paprastumas bei virtualaus mokymosi teikiama galimybė mokytis pačių pasirinktu laiku, patogioje vietoje.

Nuomonės apie kurso trūkumus pasiskirstė įvairiai. 18 studentų atsakė, jog kurse nėra trūkumų, 1 studentas pasigedo galimybės pretenduoti į stipendijas.

Taip pat pasidomėta, ar studentai gali vertinti kursą esant naudingu, įdomiu ir išskirtiniu. 18 studentų (94,7%) sutinka su minėtuoju kurso apibūdinimu, 1 studentas (5,3%) nesutinka.

Kitu klausimu buvo studentų paklausta, ar visų tipų kalbines įgūdžius lavinančias užduotis, jų nuomone, galima rasti virtualiame kurse. 17 studentų atsakė teigiamai, 2 pateikė neigiamą atsakymą.

Galiausiai studentų buvo paklausta, ar norėtų išmokti kitų užsienio kalbų naudojantis tokio tipo kursu kaip „Lithuanian language for foreign students“. Visi studentai į šį klausimą pateikė teigiamą atsakymą.

Atlikta apklausa parodė, jog į KTU iš užsienio šalių atvykę studentai lietuvių kalbos mokymąsi virtualiu būdu laiko priimtiniu kalbos mokymosi būdu. Tikslingai mokymosi procese parinktos užduotys, jų paprastumas bei atitinkamų IT priemonių panaudojimas, studentams yra efektyvus žinių įgijimo būdas, nepaisant keleto trūkumų, kurie buvo įvardinti.