

# Turinys

1	Įvadas .....	3
2	Šiuolaikinės kompiuterinės vertimo sistemos analizė.....	6
2.1	Kompiuterinio vertimo sistemų vystymo ir diegimo tendencijos .....	6
2.2	Kompiuterinis vertimas į lietuvių kalbą.....	7
2.3	Sistemos architektūriniai sprendimai .....	8
2.3.1	Sprendimų lentos modelis .....	8
2.3.2	Tiesioginis .....	12
2.3.3	Transformacinis.....	14
2.4	Vertimo žodynai-leksikonai .....	15
2.5	Kompiuterinio vertimo varyklės .....	16
2.5.1	Pažodinis .....	16
2.5.2	Interlingua .....	16
2.6	Kompiuterizuoto vertimo proceso funkcinė struktūra .....	23
2.7	Gramatinis šaltinio teksto nagrinėjimas.....	26
2.7.1	Segmentinis sakinio analizės būdas .....	26
2.7.2	Automatinis sakinio analizės būdas .....	27
2.7.3	Taisyklių nustatymo sakinio analizės būdas .....	28
2.7.4	Hibridinis sakinio analizės būdas.....	29
3	KV iš anglų į lietuvių kalbą sistemos projektas .....	31
3.1	Žodynų pildymas.....	33
3.2	Gramatinis sakinių nagrinėjimas.....	36
3.3	Vertimai iš anglų į lietuvių kalbą.....	39
4	Ekspirimentinis KV sistemos tyrimas .....	41
4.1	Ekspirimentai su anglų-lietuvių kalbos sistema .....	41
4.2	Palyginamieji ekspirimentai su anglų-rusų-anglų PROMTX, laalvera ir nauja „EiSee“ vertimo sistema .....	42
4.3	Ekspirimentų rezultatai ir jų analizė.....	<b>Error! Bookmark not defined.</b>
5	Išvados.....	46
5.1	Kas padaryta.....	<b>Error! Bookmark not defined.</b>
5.2	Ekspirimentas ir jo rezultatai.....	<b>Error! Bookmark not defined.</b>
5.3	Kur gali būti pritaikyta .....	<b>Error! Bookmark not defined.</b>
6	Literatūros sąrašas.....	48
7	Santrauka.....	49
8	Terminų ir santrumpų žodynas.....	50
9	Priedai.....	52
9.1	KV rezultatas su gramatine sakinio analize .....	52
9.2	Kompiuterinis vertimas naudojantis laalvera programa.....	55

# 1 Įvadas

Kompiuterinio vertimo pirmieji tikslai buvo sukurti visiškai automatinio vertimo sistemas, kurios nereikalautų žmogaus pagalbos ir savo vertimo kokybe prilygtu žmogaus verstam tekstui. Kaip vėliau paaiškėjo, kad tai kol kas neįmanoma, nes tokių sistemų kokybė labai žema ir ji netenkina vartotojų poreikių. Žmogus turi peržiūrėti visiškai automatinių KV sistemų rezultatus, kad pasiektu reikiamos kokybės. Visiškai automatinės KV sistemos nėra bevertės. Net ir neredaguotas KV tekstas gali būti naudingas, pavyzdžiui tiems, kuriems reikalinga suprasti nežinomo teksto esmę per trumpiausią laiką ir mažiausiom išlaidom. Nors ilgą laiką buvo siekiama kokybės lygintinos su žmogaus vertimo kokybe, tačiau nepasiekus reikiamo tikslo, buvo pradėtos taikyti KV sistemos su žmogaus pagalba. Tai ekonomiškai efektyvi priemonė daugianacionalinėms korporacijoms ir daugiatautėms organizacijoms (tokioms kaip Europos Sąjunga). KV sistemos verčia ganėtinai grubiai, todėl reikalingas teksto redagavimas. Teksto redagavimas gali būti brangus vertimo būdas. Norint pasiekti didesnės kokybės ir ekonomiškai naudingo sprendimo, imta taikyti valdomas kalbas, sumažinant ar net pašalinant leksines dviprasmybės ir sudėtingų sakinių supaprastinimas. Taip vertimo procesas yra susiejamas su techninės dokumentacijos rašymu, kuris ateityje sumažins vertimo laiką ir kaštus.

Pasaulyje žinomi du pagrindiniai kompiuterinio vertimo sistemų tipai. Visiškai automatinės, kurios bando versti visą tekstą be vartotojo įsikišimo, viso vertimo proceso metu. Kadangi tokių sistemų vertimo rezultatas yra gana žemos kokybės, tai reikalingas žmogaus įsikišimas. Kitos rūšies KV sistemos turi įvairiausių vertėjo pagalbos įrankius, kurie teikia lingvistinę pagalbą vartotojui, daugiausia žodynais ir gramatine, o taip pat ir tai kas yra vadinama valdoma kalba, t.y. prieš tai jau verstų tekstų duomenų bazė, kad galėtų pagelbėti vertėjui.

Lietuvos integravimasis į NATO ir ES skatina plėtoti ir nacionalinę kultūrą, o taip pat ir savo kalbą, nes šių organizacijų dokumentai bus išversti į kalbas šalių sudarančių šias organizacijas. Taigi gali išlikti tik moderni ir į šiuolaikinius bendravimo poreikius orientuota kalba. Dar nei viena mažos šalies kalba neišvengė poveikio didesniųjų. Šiais globalizacijos laikais, svarbus filologijos specialistų tikslas – kiek įmanoma, išlaikyti gimtąją kalbą autentiška. Lingvistikos specialistai, susidomėję praktiniu kalbos panaudojimu, ieško naujų būdų kaip tai įgyvendinti.

Nepaisant sparčios informacinių technologijų plėtros, daugiakalbės tinklo ir kompiuterių sistemų technologijos nėra sukurtos lietuvių kalbai. Dėl anglų kalbos žinių stokos, daugiau kaip 65% lietuvių yra atskirti nuo elektroninės komercijos, neskaitant galimybių prieiti prie interneto. Lietuvių

kalba naudojama tik ten kur visuomet bus reikalinga, specializuotose vyriausybių Įstaigų programose, mokyklų mokomosiose programose, kai tuo tarpu dauguma taikomųjų programų prieinamos kitomis, dažniausiai anglų kalba. Todėl žmonėms su žemesniu išsilavinimu, kaimo gyventojai, seneliai ir kt. atsiduria pavojuje būti atkirstiems nuo plėtojimosi.

Internetas puiki priemonė bendrauti su visu pasauliu. Kai kurioms Įmonėms gal net būtinybė, kad galėtų plėtoti savo verslą. Augant vartotojų poreikiams bendrauti su pasauliu pigiausiomis priemonėmis. Internete bendraujama ne viena, nors dažniausiai anglų bet ir kitomis kalbomis. Jau keletą metų internete teikiama teksto ir interneto puslapių vertimas iš anglų Į ispanų, prancūzų, vokiečių, italų ir portugalų, o taip pat ir iš ispanų, prancūzų ir vokiečių Į anglų kalbą. Absoliutus tikslumas yra problema, kai verčia programas, o ne žmonės. Bet jei naudojami bendram medžiagos supratimui, tuomet gali būti efektyvi pagalbos sistema mokymuisi ar bendravimui.

Idėjos apie visiškai automatinės sistemos sukūrimą, galinčios versti tekstus savo kokybe prilygstantys žmogaus vertimo kokybei, tampa vis mažiau populiaros, nes kiekvienam profesionaliam vertėjui nepriimtinos tokios KV sistemos, dėl pernelyg žemos kokybės. Vadinasi dėmesys turėtų būti nukreiptas Į tokias sistemas, kurios leistų prisitaikyti sistemą sau pačiam. Dabar orientuojamasi Į atvirojo kodo sistemas, kurios savo pagalbos priemonėmis vartotojui leistu lanksčiai prisitaikyti prie kiekvieno iš jų.

Lietuvoje KV srityje, dar ne daug tenuveikta. Sukurta virtualaus žodyno sistema (2002m.), sukurta M. Žemaičio LAAL anglų-lietuvių kalbos vertimo programa naudojanti sprendimų lentos modelį (2002m.).

Mano planas – sukurti sistemą galinčią versti iš anglų Į lietuvių kalbą, naudojantis valdomos kalbos modelių, naudojant sintaksinę ir semantinę sakinio analizę. Šiam tikslui panaudosime, labiausiai mūsų poreikius atitinkančią sistemą, šaltinio teksto analizei atlikti. Mano siūlomoje kompiuterinio vertimo programoje nenaudosime „Interlingua“, nes tokių sistemų kokybė labai prasta, todėl panaudosime sintaksės ir semantikos analizę, kad pasiektumėm geresnių vertimo rezultatų. Diegsime valdomos kalbos modelį su pagalbos priemonėmis vartotojui.

KV sistemas kuria didelės Įmonės ir organizacijos, skiriančios milijonus eurų vienai kalbai. Pigios ir prastos kokybės sistemos užsakovams kainuoja apie 1.5-2,5 mln. eurų vienos kalbos kūrimui (promptX - rusų) ir trunka apie porą metų, o vartotojui programa kainuoja 250 eurų (anglų-rusų, rusų-anglų) ir po 100 eurų už kiekvieną papildomą specializuotą žodyną. Geros kokybės vertimo sistemos

kainuoja apie 20 mln. eurų ir dirbama 10-15 metų ir daugiau. Prancūzijos vyriausybės finansuotas projektas eiliniams vartotojams yra prieinamas nemokamai.

Valdomos kalbos sąvoka kalbų inžinerijoje pagal EAMT (European Association for Machine Translation, 2003 m.) koncepciją, suprantama kaip išskirtinis „*bendrinės kalbos poaibis*“, kuris nustatomas pagal tam tikrus principus. Jį gali sudaryti dvi „išvestinių kalbų“ kategorijos: *pirmoji*, skiriama vartotojams, kurie bendrauja ne savo gimtąja kalba, *antroji*, gi, orientuota tiems, kas nori taikyti kompiuterines (arba kitokias informacines) kalbų inžinerijos technologijas praktikoje. Pastaroji siejama su ribotų arba specializuotų tekstų transformavimu iš vienos kalbos (toliau šaltinio) į kitą (tikslo) kalbą kompiuterizuoto vertimo (KV) sistemose. Toks apribojimas atsispindi žodynų (lexicons) sudarymo ir naudojimo procedūrose, yra daugiau skirtas specializuotų tekstų analizei ir remiasi supaprastintomis lingvistinėmis taisyklėmis arba siauresne jų interpretacija.

Suprantama, kad perspektyvios kompiuterinės tekstų analizės technologijos turi palaikyti *atviros sistemos* modelį, kuris leistų vartotojui laisvai manipuliuoti ne tik pažangiomis informacijos apdorojimo galimybėmis, bet ir atitiktų jo pageidavimus. Mūsų siūlomas KV sistemos modelis tenkina visas atviros sistemos sąlygas ir numatomas realizuoti per atitinkamus programinius modulius: virtualųjį žodyną, specializuotus žinių šaltinius, vertimo valdymo bloką, dialoginės vertimo terpės realizavimo principus, sintaksės ir semantikos analizę.

Valdomos kalbos modelis buvo pristatytas 2004 m. sausio mėn. KTU organizuotoje konferencijoje „*Informacinės technologijos kalbų inžinerijoje*“. A. Sipavičius ir B. Tamulynas „*Valdomojo kompiuterinio vertimo technologijų galimybių tyrimas*“.

## 2 Šiuolaikinės kompiuterinės vertimo sistemos analizė

### 2.1 Kompiuterinio vertimo sistemų vystymo ir diegimo tendencijos

Kompiuterinio vertimo technologijų kūrimo pradžioje (1960 m.), pagrindiniu tikslu buvo laikoma visiškai automatizuota sistema su aukščiausia vertimo kokybe, lygintina su žmogaus vertimu. Tokios sistemos pagalba galėtume tuojau pat publikuoti norimą tekstą, nes jis nereikalauja žmogaus įsikišimo. Tačiau tokios sistemos sukurti nepavyko. Rezultatai parodė, kad publikuojamos kokybės negalima pasiekti visiškai automatinio vertimo sistemomis, todėl būtina žmogaus pagalba, tokiomis priemonėmis kaip: išversto teksto (tikslo kalbos) peržiūra, ar verčiamo teksto (šaltinio kalbos) kontroliavimas (supaprastinimas, gramatikos reguliavimas ar žodynų valdymas). Net ir žemesnės nei puikios vertimo kokybės sistemos gali būti naudingos informacijos supratimo tikslams. Dabar pagrindinė automatinio vertimo idėja yra ta, kad ne visada būtina aukščiausia kokybė, o svarbu suprasti tik esmę.

Kaip J. Hutchins teigė “*Introductory speech at the MT Summit* ” konferencijos įžanginėje kalboje, jog mes turime kurti sistemas su pagalbos funkcijomis ar įrankiais dvikalbiam ar daugiakalbiam bendravimui. Esamoms sistemoms priešpastatoma KV sistemos ir vertėjo komunikavimo pagalbos priemonės. Išskirtinas dėmesys turėtų būti tyrimui ir kūrimui įrankių, bendravimui tarp skirtingų kalbų, kai yra labai platus ratas skirtingų kriterijų sprendžiančių ar reikiami poreikiai buvo patenkinti.

Pagrindiniai sistemų ir pagalbos priemonių tipai:

1. Tradicinis KV su tam tikru apdorojimu, kai rezultato kokybės gerinimas vyksta valdant šaltinio kalbos tekstą (priešlaikinis redagavimas ar panaudojus valdomas kalbas) arba kai vartotojas redaguoja gautą tekstą. Naudojamas vertingos informacijos (dokumentacijoms ir kt.) vertimui, kai reikalaujama geros kokybės ir reikalingas ekonomiškąs sprendimas. Toks KV tekstas yra kompiuterio verstas juodraštinis variantas (*computer-produced draft translation*).
2. Tradicinės kompiuterinio vertimo pagalbos priemonės, dažniausiai naudojamas profesionalių vartotojų, kurie valdo dvikalbius žodynus, reguliuoja terminologijos pritaikymą, vartotojo atmininės naudojimas. Skirtos įdiegti tam tikrose tarnybinėse stotyse.
3. Priemonės dokumentų ar informacijos asimiliavimui kitomis kalbomis. Tai tradicinis KV sistemos panaudojimas dokumentų filtravimo paieškoje, o dabar dar ir tinklo puslapių naršymui.
4. Priemonės teksto vertimui kita kalba. Tai naujas būdas ir naudojamas

- a. sistemose (ar pagalbos priemonėse) daugiakalbiame techninės dokumentacijos vertimui,
  - b. naudojamas KV sistemose, kai nėra tiksliai žinoma šaltinio kalba,
  - c. užsienio kalbų autorinėse programose (pvz.: padėti Japonų autoriams parašyti anglišką konferencijos straipsnį).
5. Priemonės padedančios suprasti esmę, kai nepagrindinis tikslas yra tikslus vertimas, bet esmės perteikimas. Toks vertimo būdas gali būti naudojamas TV programų pavadinimams ar subtitrams, skelbimams (geležinkelio, autobusų stotyse), policijos pranešimuose ir t.t.
  6. Priemonės informacijos gavimui per tarpinę kalbą t.y. gauti informaciją kitomis kalbomis.
  7. Pagalbos priemonės palaikančios tarpinės kalbos informacijos apsikeitimui. Skirtas elektroniniuose laiškuose, kitoje neoficialioje spaudoje.

Išskyrus 1 ir 2 tipą, vertimas yra grubus. Pirmas ir antras tipas reikalauja tikslaus vertimo. Čia tik priemonės kuriomis gali naudotis žmonės. Jos efektyviau leidžia mums bendrauti, bet ne mus pakeisti.

Ir toliau didės geros kokybės KV sistemų poreikis. O dabar reiktų sutelkti dėmesį dviejų tipų pagalbos priemonių plėtotei:

- a) vertimo, kuris gali būti redaguojamas, kol pasiekama reikiama kokybė,
- b) vertimui perteikiančiam teksto (žinutės) esmę.

## **2.2 Kompiuterinis vertimas į lietuvių kalbą**

KV sistemose ypač opią problemą sudaro sinoniminis ir kontekstinis daugiaprasmiškumas. Be to, kiekviena kalba turi tik jai būdingą vienokią ar kitokią žodžių tvarką sakinyje, kai kurie žodžiai neturi atitikmens tikslinėje kalboje ir panašiai. Dėl šių ir kitų priežasčių vertimas interpretuojamas kaip tam tikras kūrybinis aktas. Nepaisant minėtų išskirtinių atvejų ir naudojant sudėtingesnius originalaus ir tikslinio teksto semantinės analizės programinius modulius, vertimo procesas su tam tikromis išlygomis gali būti kompiuterizuotas.

Verčiant iš anglų į lietuvių kalbą didelė problema – sinoniminis daugiaprasmiškumas, nors tai ne vienintelė problema, tačiau viena svarbiausių. Tam kad sakinyb būtų teisingas, būtina:

- Parinkti ir suderinti gramatinės žodžių formas.
- Suderinti gimines.
- Skaičius.
- Laipsnius.

- Pritaikyti linksnių ir asmenų galūnes.
- Nustatyti bendrines ir tikrinės formas.
- Nuspręsti kuri veiksmažodžio nuosaka tinkamesnė, nors neretai veiksmažodžių nuosakos gali būti vartojamos viena vietoje kitos.
- Sangražinių veiksmažodžių ir iš veiksmažodžio padarytų daiktavardžių naudojimas.  
Nustatyti kada sangražinės formos negalima vartoti.

Teisingas sakinio sudarymas svarbi problema, bet ir ją išsprendus lieka dar daug problemų, tokių kaip kontekstinio daugiaprasmiškumo sprendimas t.y. dviejuose skirtinguose tekstuose esantis vienas ir tas pats sakinyš privalo būti verčiamas skirtingai, kad neprarastumėm teksto turinio prasmės. Kontekstinio daugiaprasmiškumo gerų problemos sprendimų nėra. Šiai problemai spręsti naudojami euristiniai metodai pagerinantys teksto prasmę, nors neretai pasitaiko ir atvirkščiai.

Vertimas į rusų kalbą kiek kitoks nei į lietuvių, nors sakinio derinimas panašus kaip ir lietuvių kalba, reikia derinti laikus, giminės, skaičių ir laipsnius, tačiau rusų kalba gera tuo, kad neiškyla problemų kaip sangražinių formų naudojimas. Rusų kalbos lingvistai naujadarus daro dažniausiai iš anglišku žodžių (pvz. angliškai *hacker* – rusiškai *haker*), tuo tarpu lietuvių kalbos lingvistai žodį „*hacher*“ siūlo versti kaip programisiusas. Lietuvių kalboje dažnai keičiami naujadarų pavadinimai, taip pat nekelia KV sistemų kokybės.

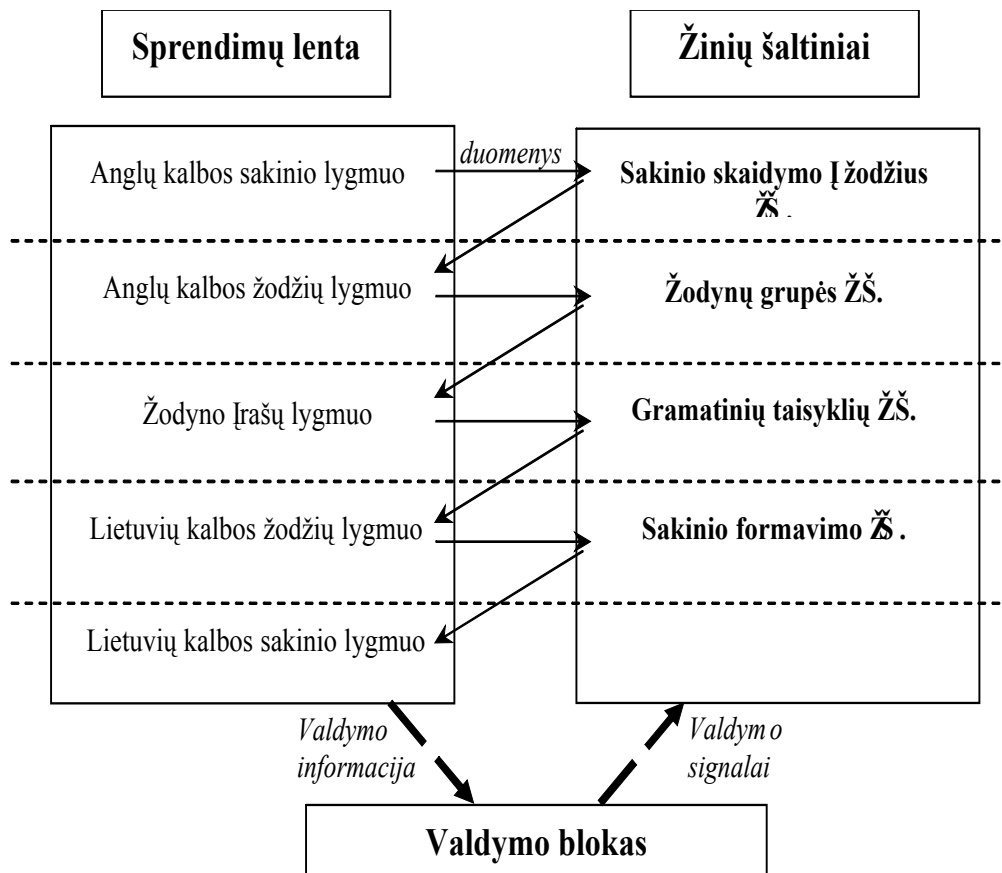
Iš anglų verčiant dažniausiai pasitaiko asmens nustatymo problema, nes anglų kalba „*you*“ reiškia ir „*jūs*“ ir „*tu*“ (*rusiškai* „*vi*“ ir „*ti*“).

## 2.3 Sistemos architektūriniai sprendimai

### 2.3.1 Sprendimų lentos modelis

Valdomojo KV sistemos architektūros prototipu buvo pasirinkta sprendimų lentos paradigma [13,14,15] (žr. 1 pav.). Ją sudaro sprendimų lenta (bendras veiksmų laukas, kuris skaidomas į dvi dalis išeities ir tikslinės kalbos tekstams saugoti), žinių šaltiniai (aktyvūs tekstų analizės ir vertimo varyklės moduliai, kurie gali tiesiogiai „veikti“ bendroje atminties srityje) ir valdymo blokas kuris stebi, aktyvina ir valdo vertimo procesą, jo tikslas kokybiškai palaikyti sąsają su vartotoju vertimo procese. Tokiu būdu vertimo eiga yra ne vienkartinis žingsnis, kurio rezultatas yra galutinis teksto vertimas, bet savotiškas daug ciklų turintis procesas. Dėl to vartotojas turi galimybę vėl ir vėl grįžti prie vertimo, jį keisti ir tobulinti, pasinaudodamas papildomais žinių šaltiniais. Patį vartotoją sistema interpretuoja kaip tam tikrą žinių šaltinį, kuris turi jam priskirtas specifines galimybes. KV sistema pasižymi adaptyvumo savybe, nes žinių šaltiniai, būdami nepriklausomi, gali įsijungti į vertimo

procesą ir veikti pagal valdymo komandą, kai jiems veiksmų lauke sukuriama palanki situacija, atitinkanti jų žinių lygį.



1 pav. Sprendimų lentos KV sistemos architektūra

**Sprendimų lenta.** Lentą galima interpretuoti kaip didelės apimties duomenų bazę ar atminties sritį, kurios turinys yra pasiekiamas kiekvienam žinių šaltiniui ir tik jie gali daryti pakeitimus šioje srityje. Bazę sudaro hierarchiniai sluoksniai, kuriuose atsispindi skirtingi uždavinio abstrakcijos lygiai. Šioje bazėje gali būti saugomi įvairūs objektai, kurių kiekvienas turi specifinius atributus. Tarpusavyje objektai siejami ryšio atributais. Žemiausiame lentos sluoksnyje saugomi duomenys, aukščiausiame – problemos sprendinys.

**Žinių šaltiniai (ŽŠ).** Tai išskaidytos probleminės srities žinios, kurias galima interpretuoti kaip atskirus specialistus (ekspertus). Kiekvienam žinių šaltiniui priskiriamos dvi sprendimo priėmimo fazės: išankstinių sąlygų įvertinimas ir elgsenos (veiksmų) fazė. Žinių šaltinio funkcinė sprendinio priėmimo paskirtis yra veikti pagal produkcinę “jeigu – tai” tipo taisyklę, kurios sąlyga – išankstinės sąlygos, o kūnas – vykdymo blokas. Žinių šaltiniai gali būti realizuojami įvairiai: naudojant loginių



taisyklių ar procedūrų rinkinį, semantinius ar neuroninius tinklus [14] ir pan. Tai savotiška specializuota ekspertinė sistema, turinti atskirą žinių bazę bei sprendimo išvedimo mechanizmą.

**Valdymas.** Pagrindinė valdymo funkcija – išrinkti naudingiausią sprendimo procesui žinių šaltinį ir jį aktyvinti. Vienu metu gali būti tenkinamos kelių žinių šaltinių aktyvumo sąlygos, t.y. pretendentų gali būti keli ir tokiu būdu turėsime veiksmo pasirinkimo konfliktą. Kiekvienas žinių šaltinis pateikia savo pasiūlymus, kuriuose nurodo, kokius pakeitimus sprendimo lentoje ruošiasi atlikti. Pagal šiuos pasiūlymus valdymo posistemė išsirenka, kurį žinių šaltinį aktyvinti. Paprasčiausiu atveju valdymo mechanizmas aktyvina tą žinių šaltinį, kuris anksčiausiai pateikė pasiūlymą. Tokia valdymo strategija vadinama oportunistine. Ji reglamentuoja nuoseklų sistemos valdymą, nes sprendimo priėmimo žingsniai priklauso nuo esamos situacijos sprendimo lentoje [13]. Kito tipo valdymo strategija – planuojamoji, kai valdantieji konfliktus sprendimai vykdomi pagal tam tikrą planą.

Sistemos darbo efektyvumą galima padidinti vienu metu nagrinėjant kelias hipotezes ar priimant sprendimus keliuose hierarchijos lygiuose. Todėl, didėjant kompiuterių spartai, vystomos lygiagrečios sprendimų lentos modelio architektūros [13].

**Sakinio skaidymas į žodžius.** Šaltinio teksto skirstymas (*parsing*) [2,8,9,11] yra esminis ir pirmaeilis uždavinys, kuris turi būti sprendžiamas KV vertimo įžangoje. Skirstymo programos rezultatas yra šaltinio kalbos žodžiai, kurie turi tam tikrą struktūrinį aprašą pagal priimtas tos kalbos gramatikos taisykles. Intelektualiosios skirstytojo programos „sugeba“ įvertinti kalbos semantinius atributus ir tokiu būdu palengvina sakinio vertimą, išverstas tekstas tampa semantiškai adekvatus įeities tekstui.

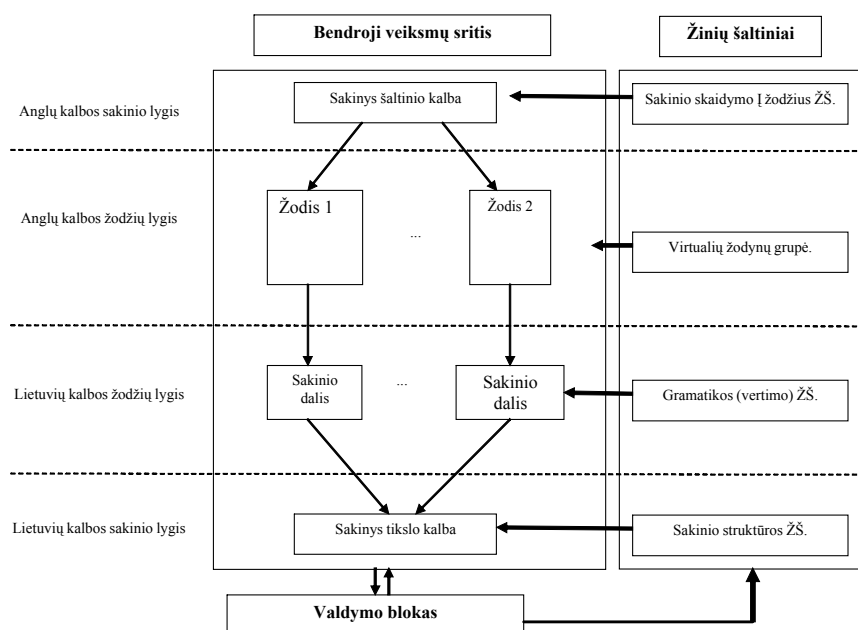
**Kompiuterinio vertimo žodynas.** Žodynai būtini KV sistemoms. Be jų KV sunkiai įsivaizduojamas. Žodynai dažniausiai būna daugiaprasmiai. Vienas žodis gali turėti daug skirtingų jo vertimo variantų, skirtingomis kalbos dalimis. Kuo žodynas pilnesnis, tuo mažiau reikės vartotojui juos įvedinėti, žinoma, jei yra tokia galimybė. Žodynai būna: dvikalbiai ar daugiakalbiai, žodžių, terminų ar frazių. Jais naudojantis galimas KV.

Vertimo kokybę galima pagerinti išnaudojant žodyno virtualumo savybę, t.y. sudarant galimybę nuolat jį papildyti naujais žodžiais, terminais, frazėmis ir frazeologizmais. Tačiau tai veda prie sudėtingesnės teksto analizės ir todėl tenka apsiriboti žodžių junginiais, kurių vienus racionaliau saugoti bendrame žodyne, o kitus specializuotame, susiejant juos per žodžius, įeinančius į sakinį.

**Kitų kalbos žinių resursų panaudojimas.** Šiuo metu, kai kompiuterių atminties resursai ženkliai išaugo, tikslinga kurti moksliają vertimo sistemą, kurios žinių bazėje būtų saugomi vertimo pavyzdžiai. Tokiu būdu sistema, suradusi teksto fragmentą ir pažymėtą kaip tam tikrą žodžių grupę, terminą ar frazę, imtų iš atminties atitinkančią frazę, kuri anksčiau jau buvo išversta, ir taip sparčiau

leistų generuoti teksto vertimą. Panašus principas yra siūlomas naujos kartos KV sistemų architektūroje vadinamas „*vertėjo atmintine*“ – *translation memory* .

**Tekstynų naudojimas.** Tekstynų panaudojimas vertimų praktikoje, kaip, beje, ir kitose kalbų inžinerijos srityse yra nauja ir pažangi technologija. KV sistema kelia kiek didesnius reikalavimus atitinkamai tekstyno duomenų bazei ir jos pritaikomumui. Pirma, toks tekstynas turi būti daugiakalbis. Antra, turi būti garantuota pakankamai tobula jo sąsaja su KV sistemos paieškos ir teksto analizės posistemėmis. Trečia, probleminės srities tekstų rinkinys turėtų būti pakankamai informatyvus ir reprezentatyvus.



2 pav. Kompiuterizuotos vertimo sistemos sprendimo lentos modelio veikimo schema

Vertimo varyklė siūlomo modelio kontekste gali būti realizuota vienu ar keliais žinių šaltiniais, papildant juos autorine tekstų koregavimo galimybe vartotojo lange. Išvardytus principinius varyklių privalumus ir trūkumus sunku vertinti vienareikšmiškai. Gali pasirodyti, kad pažodinis vertimas yra netobulas. Tačiau yra žinoma nemažai pažodinio KV sistemų, kurios sėkmingai platinamos. Iš tikrųjų, grynai pažodinio vertimo sistemų praktiškai ir nėra, nes kiekviena KV sistema realizuoja atitinkamas metodų kombinacijas. Pavyzdžiui, verčiant *perstatymo* metodu dažnai tenka pasinaudoti ir *tiesioginiu* vertimu, nes sakinio analizė ne visada gali būti patenkinama. Todėl, atsižvelgiant į tai, kokio rezultato mes siekiame, galime pasirinkti vieną iš minėtų vertimo principų.

Su sinoniminio ir kontekstinio daugiareikšmiškumo problemomis lanksčiau susidoroja KV sistemos, kurios naudoja kalbos žinių šaltinius, kaip nepriklausomus programinius vienetus ar komponentus (Tamulynas, 2001). Kaip antai: teksto ir sakinio analizės, žodžių analizės moduliai, gramatinės ir semantinės analizės komponentai, žinių gavyba iš tekstyno ir pan. Teksto analizės arba

skirstymo šaltinis išskiria sakinius, o juose žodžius. Žodžių analizės šaltinis pateikia informaciją apie žodžių semantines, sintaksines savybes ir ieško frazės ar žodžių atitikmens žodynuose. Gramatinės analizės modulis normuoja žodžių tvarką sakiniuose ir pan. Jo tikslas sutvarkyti linksnius, laikus ir teisingai surikiuoti žodžius sakinyje. Sintaksinių sakinio grupių sudarymas ir jų panaudojimas gramatinės analizės komponente yra vienas iš būdų, kuris leidžia pasiekti geresnę kompiuterinio vertimo kokybę (Žemaitis, 2002).

Taikant aprašytą sprendimų lentos modelį ir atsižvelgiant į kompiuterinio vertimo proceso aiškiai išreikštą hierarchinę struktūrą, galima skirti atitinkamas žinių šaltinių grupes:

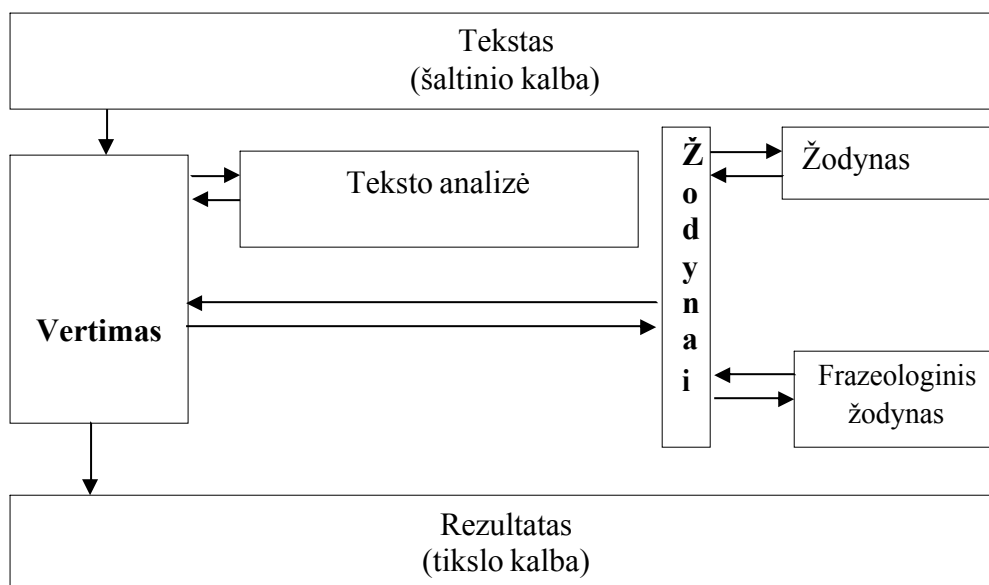
- virtualiųjų žodynų grupė: bendros paskirties virtualus žodynas, frazeologinis virtualus žodynas, teminiai (specialieji) virtualūs žodynai, specialieji daugiakalbiai tekstynai;
- sintaksinė sakinio analizė;
- gramatinės analizės ŽŠ grupė;
- pagalbiniai ŽŠ, kurie „išmano“ ir yra atsakingi už vertimo konteksto įvertinimą, originalo ir vertalo semantinį adekvatumą bei kt.;
- probleminių žinių gavyba iš specialiųjų tekstynų ir jų panaudojimas tekstų semantinei erdvei formuoti.

Kompiuterinės vertimo sistemos vartotojai (vertėjai) bei korektoriai (ekspertai) sprendimų lentos modelyje gali būti interpretuojami kaip specialieji ŽŠ, jei jie yra aktyvūs dialogo dalyviai ir jų veiksmai turi įtakos kompiuterinio vertimo proceso kokybei. Elementarus konceptualiojo kompiuterinio vertimo sistemos hierarchinis modelio variantas su specifikuotais ŽŠ vaizduojamas 2 paveikslėlyje. Valdymo bloko strategija šiuo atveju yra artimesnė veiksmų planavimui ir grįžtamojo ryšio palaikymui. Pagrindinės valdymo bloko veiksmų vykdymo funkcijos: žinių šaltinių veiksmų koordinavimas, tekstų adekvatumo kokybės kriterijų įvertinimas bei vertimo variantų skaičiaus reguliavimas.

### **2.3.2 Tiesioginis**

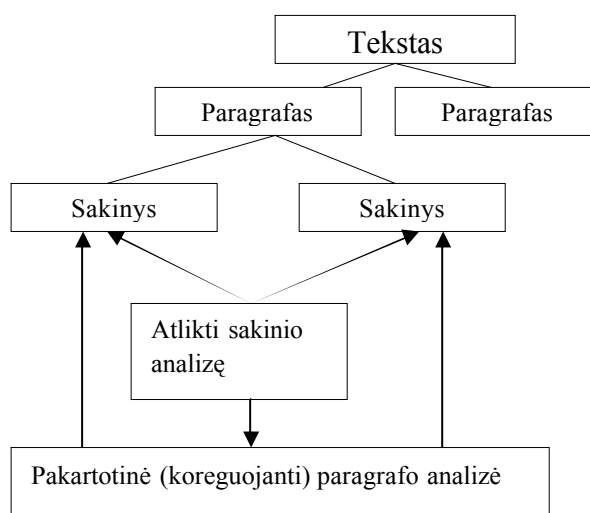
Tiesioginis – toks modelis, kuris yra itin paprastas. Jį sudaro vertimo blokas, kuris iš šaltinio kalbos verčia į tikslo kalbą naudodamasis teksto analize ir žodynais. Teksto analizė atlieka viso teksto suskaidymą į smulkius modulius (paragrafus, sakinius, žodžių grupes ir žodžius). Tiesioginis modelis retai taikomas, nes tokių sistemų kokybė labai prasta ir beveik niekam netinka. Tai greičiau skirtas pažodinio vertimo sistemoms. Nors kaip minėjau tokia architektūra neleidžia pasiekti kokių nors kokybės aukštumų, tačiau pasaulyje yra pažodinių vertimo sistemų, kurios sėkmingai platinamos. Tokios sistemos nyksta ir tikriausiai išnyks atsiradus efektingesnėms sistemoms, kurios labiau tenkins vartotojo poreikius už tą pačią kainą. Vertimo modelis naudojasi dviem: pagrindiniu ir frazeologiniu

žodynu. Pagal teksto analizės sudarytas žodžių grupes bandoma surasti atitikmenų frazeologiniame žodyne, neradus žodžių grupės verčiamos pažodžiui.



3 pav. Tiesioginio modelio schema

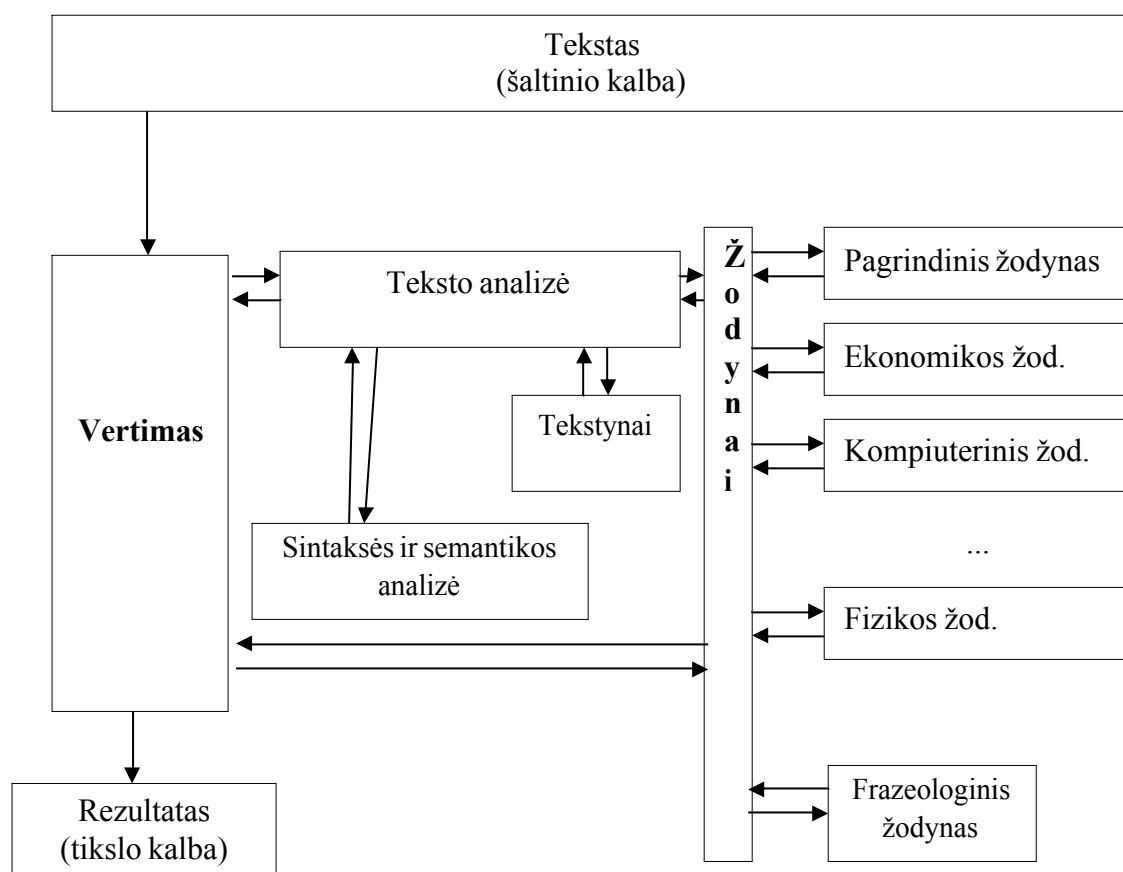
Teksto analizė. Tekstai pagal savo turinį skirstomi į straipsnius. Straipsniuose nauja mintis pradedama nauja pastraipa. Šis skirstymas leidžia manyti, kad pastraipoje aprašoma atskira mintis, pastraipa buvo rašoma to paties rašytojo, kaip tik dėl to galimas teisingesnis vertimas. Atlikus pastraipoje kiekvieno sakinio analizę, reiktu atlikti pakartotinę kiekvieno sakinio analizę, kad pritaikyti kiekvieną sakinį prie pastraipos t.y. parinkti vieną teminį žodyną labiausiai tinkantį visai pastraipai, suvienodinti terminus, kad keliuose sakiniuose vienas ir tas pats daiktas turėtų tą patį vertimą, o ne sinoniminį vertimą.



4 pav. Bendras teksto analizės būdas

### 2.3.3 Transformacinis

Transformacinis modelis – tiesioginio modelio atmaina su papildomomis priemonėmis. Nuo tiesioginio modelio skiriasi tuo, jog vertimo modulis ne tiesiogiai siejamas su žodynais, o teksto analizės modulis turi priėjimą prie žodinių, todėl žodžių grupės sudaromos tikslingai (tinkamai parenkami terminai). Čia taip pat naudojami specializuoti terminų žodynai, kuriais gali būti naudojamosi priklausomai nuo to, kokia tematika yra verčiamas tekstas. Bet pagrindinis ir tikriausiai svarbiausias yra sintaksės ir semantikos analizės blokas. Sintaksės ir semantikos nagrinėjimas pagerina vertimo kokybę. Transformacinis modelis gali pagerinti vertimo kokybę, bet jis nėra tobulas, nes neprisitaiko vertimo proceso metu, todėl jo vertimo kokybė nebus tokia gera kaip sprendimų lentos modelyje.



5 pav. Bendra transformacinio veikimo schema

## 2.4 Vertimo žodynai-leksikonai

**Kompiuterinio vertimo žodynas.** Iš kitų kalbų lietuvių kalbos žodžiai išsiskiria sandaros ir sintaksės sudėtingumu bei prasmių gausumu. Kompiuterinei sistemai, kuri verčia tekstą iš originalo kalbos į lietuvių (arba priešingai), reikalingas dvikalbis žodynas, kuriame žodžiai būtų pateikiami su visais morfologiniais jų kaitybos atributais ir jų sudarymo taisyklėmis. Be to, parenkant žodyno struktūrą, įvertinama, kokias semantines funkcijas vertimo procese žodynas turi palaikyti ir kokia papildoma informacija jame turi būti saugoma. anglų-lietuvių kalbų vertimo žodyną turėtų sudaryti:

- anglų kalbos žodžiai su atitinkamais gramatiniais ir semantiniais atributais bei papildomomis nuorodomis, kokius vaidmenis jie gali atlikti sakinyje, priklausydami vienai ar kitai kalbos daliai, pvz., daiktavardis gali būti subjektu, objektu ir pan., juo gali būti konkretus, abstraktus, gyvas ar negyvas daiktas (veiksmažodžiai dažniau naudojami su tam tikro tipo daiktavardžiais, t.y. semantiškai susiejant juos su tam tikra subjektų ar objektų grupe ir pan.);
- atitinkami lietuviški žodžiai su gramatiniais atributais ir informacija apie anglų kalbos žodžio semantinius atitikmenis bei galimas gramatines bei semantines interpretacijas.

Pagal lietuvių kalbos gramatikos taisykles žodyno bazėje pakanka saugoti žodžio kamieno požymius, o pagal linksniuotėms, giminei, asmenuotėms priimtas taisykles ir kitus, vertimui naudingus atributus, atitinkamas žodžio formas galima konstruoti procedūriniu būdu, t.y. prijungiant atitinkamas žodžio dalis (galūnes, linksnius ar kitas formas). Konkreti žodžio prasmė nustatoma remiantis informacija apie galimus jo naudojimo atvejus, susietumą su kitomis kalbos dalimis, dažniausiai užimamą vietą sakinyje, kontekstą ir pan. Reikšmingesnis vaidmuo tenka tariniui (veiksmažodis), nes pagal jį sprendžiama, kaip išsidėstę kiti žodžiai ir kaip jie gali būti susieti tarpusavyje (kontekstinė priklausomybė). Ypatingą vietą nustatant žodžio linksnio atitikmenį tekste, turi prielinksniai. Žodyne saugomi visi prielinksniai su nuorodomis, kokio linksnio jie reikalauja arba pateikiamos taisyklės, kaip jie turi būti derinami su atitinkamais žodžiais, kad žodyne būtų surastas adekvatus žodžių ar jų derinių vertimas. Žodyno struktūra yra orientuota į objektus, atvira ir veikianti virtualiai. Tai reiškia, kad žodynas gali būti lengvai papildomas, prijungiant kitas kalbos dalis su joms būdingais atributais (žr. 18-22 pav.).

Vertimo kokybę galima pagerinti išnaudojant žodyno virtualumo savybę, t. y. sudarant galimybę nuolat jį papildyti naujais žodžiais, terminais, frazėmis ir frazeologizmais. Tačiau tai veda prie sudėtingesnės teksto analizės ir todėl tenka apsiriboti žodžių junginiais, kurių vienus racionaliau saugoti bendrame žodyne, o kitus specializuotame, susiejant juos per žodžius, įeinančius į sakinį.

## 2.5 Kompiuterinio vertimo varyklės

### 2.5.1 Pažodinis

Pažodinis – toks vertimas, kai kiekvienam, šaltinio kalbos žodžiui ieškomas atitikmuo tikslo kalboje. Specialistų nuomone toks vertimo būdas niekam netinkamas, nes jo kokybė labai prasta. Kaip praktika parodė, kad tokios sistemos visgi yra naudojamos, netgi už daugumą jų – reikia sumokėti. Tačiau tos sistemos nėra tokios jau ir paprastos. Dauguma pažodinio vertimo sistemų naudoja KV juodraštinio vertimo tikslinimo ir sakinio derinimo algoritmus, kurie padeda tiksliau nusakyti mintį. Tokių sistemų žodynai dažniausiai turi tik vieną reikšmę kiekvienam žodžiui, o kad įtikti platesniam vartotojų ratui, naudojami bendresni žodžiai, todėl kartais labai tiksliai ir konkrečiai nusakantis įvykius sakinyje, gali būti verčiamas gana abstrakčiai (pvz.: „*Go home*“ turėtume išversti „*Eik namo*“, bet rezultatas būna panašus „*Vyk ti namo*“). Pažodinis vertimas kartais duoda gana suprantamą vertimą, tačiau tokių sistemų era jau baigiasi ir jos pradeda nykti iš mūsų akiračio. Pagrindiniai privalumai tokių sistemų, jos labai greitos, taupo resursus (atmintį), dažniausiai turi daugiakalbius žodynus, kurie yra paprastos struktūros ir užima nedaug vietos diske, palaiko daug kalbų, todėl tinka platesniam vartotojų ratui.

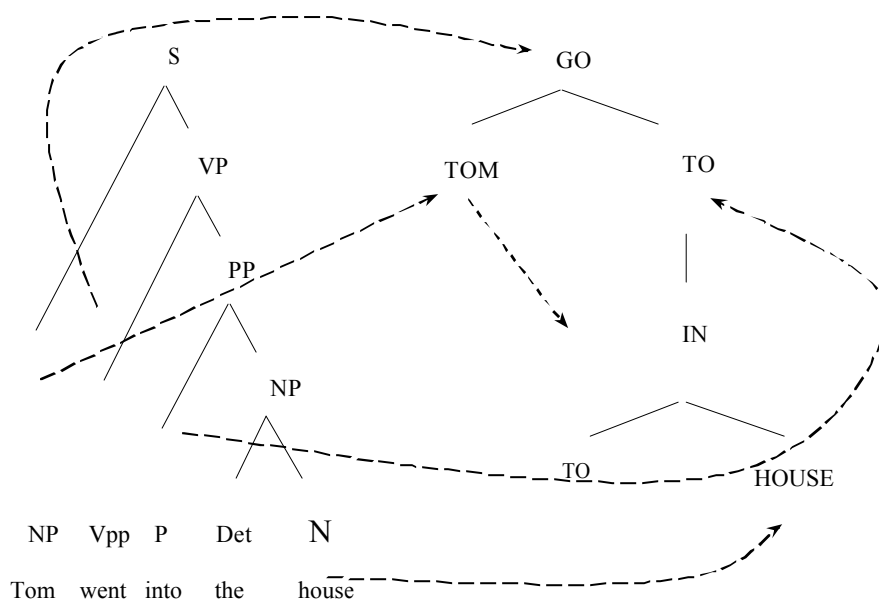
### 2.5.2 Interlingua

Universalumu ir lankstumu išsiskiriantis *Interlingua principas* vertimo kokybe nepasižymi, tačiau jis atveria formalias galimybes patogiai išreikšti teksto prasmę. Kaip pavyzdį galima paminėti leksinę konceptualiąją struktūrą (LKS), kuri buvo panaudota KV UNITRAN realizacijoje (anglų, vokiečių, ispanų kalboms). Speciali programa analizuoja sakinį, duotą kuria nors iš minėtų kalbų ir atvaizduoja jį į LKS. Iš LKS formos galima vienareikšmiškai generuoti sakinį į bet kokią kitą kalbą. 6 pav. matome schemą, kuri iliustruoja trumpo anglų kalbos sakinio transformavimo procesą į *interlingua*, o po to iš jos – į lietuvių kalbą.

Angliškai: Tom went into the house.

LKS: GO(TOM, TO(IN(HOUSE)))

Lietuviškai: Tomas įėjo į namą.



6 pav. Sakinio atvaizdavimas į LKS

LKS yra abstrakti semantinės informacijos vaizdavimo forma. Atlikus struktūrinę pradinio teksto analizę ir naudojant specialias taisykles, nustatoma priklausomybė tarp teksto ir į atitinkančio LKS tipo. Suprantama, kad LKS turi būti nepriklausoma nuo kalbos ir išpildyti principinius reikalavimus *interlinguai* atvaizduoti. Paveikslėlyje santrumpos žymi: **S** – sakinys (sentence); **N** – daiktavardis (noun) – veiksny; **V** – veiksmazodis (verb) – tarinys; **P** – prielinksnis (preposition); **Det** – pažymimasis žodis; **VP** – veiksmazodinė frazė (verb phrase) – tarinys ir frazės kartu su juo, pvz., *gerai miegu, sparčiai einu*; **PP** – prielinksninė frazė (prepositional phrase), pvz., *ant stalo, po 3 minučių*; **NP** – daiktavardinė frazė (noun phrase), veiksmi, objektai ir įvardžiai, pvz., *kiekvieną dieną, jie, namas*.

LKS tipai nustato probleminės srities esybių rūšis, kurios sudaro tam tikrą sistemą. 1 lentelėje pateikiamas galutinis visų tipų sąrašas, kuris naudojamas aprašant LKS formą. Baziniai elementai yra pritaikomi pridant lauko indikatorių, nurodantį sritį, kurioje elementas turi būti interpretuojamas. Tai yra svarbu norint nustatyti, kokio tipo yra nagrinėjamas elementas ir kartu labai naudinga informacija generuojant iš LKS į tikslo kalbą. 2 lentelėje pateikiamas išsamus bazinių elementų sąrašas. Klausiamųjų sakinių struktūra lietuvių ir anglų kalbose yra kitokia nei paprastuose sakiniuose, todėl juos atitinkanti LKS skiriasi. LKS yra rekursyvi struktūra, sudaryta iš neskaidomų dalių bei komponentų, kurie vėl gali būti LKS, t. y. vėl skaidomi kaip ir visas sakinys. Bendru atveju ši struktūra atrodytų taip:

*[Event GO<sub>Location</sub>(Thing, Path TO<sub>Location</sub>(Position IN<sub>Location</sub>(Thing, Thing)))].*



Kadangi kiekviena kalba interpretuoja tą pačią LKS skirtingais būdais, įvedus atitinkamus parametrus galima tą pačią LKS realizuoti skirtingomis semantinėmis struktūromis. Tokiu būdu parametrai papildo LKS duomenų struktūras tai kalbai būdinga informacija, kuri yra pasiekama analizės ir generavimo moduliams. Dauguma parametru iš esmės pakeičia

**1. lentelė LKS aprašyme naudojami tipai**

<i>Tipas</i>	<i>Atitikmuo</i>	<i>Paaškinimas</i>	<i>Pavyzdys</i>	<i>Kalbos ar sakinio dalis</i>
Event	Įvykis	Atitinka veiksmažodžius, atsakančius į klausimą „Kas atsitiko?“	Eiti, valgyti, sportuoti	Veiksmažodis/tarinys
State	Būsena	Veiksmažodžiai, neatsakantys į klausimą „Kas atsitiko?“	Būti, norėti	Veiksmažodis/tarinys
Position	Pozicija	Nusako situaciją su esybe abstrakčioje arba fizinėje vietoje	<i>Ant, už, po</i>	<i>Prielinksnis/ vietos apl.</i>
Path	Kelias	Nurodo konceptualią arba fizinę kryptį	Į, iš, link	Prielinksnis/ vietos apl.
Thing	Daiktas	Daiktavardžiai, kurie žymi objektus	Stalas, obuolys	Daiktavardis/ veiksnys
Property	Savybė	Kokybė, priskiriama <i>Daiktui</i>	Naujas, gražus	Būdvardis/ Pažyminys
Location	Vieta	Fizinė vieta	Namas, gatvė.	Daiktavardis/ aplinkybė
Time	Laikas	Nusako laiko trukmę, tarpinį	Po 3 val., vakar	Laiko aplinkybė
Manner	Būdas	Kokybė, priskiriama <i>Įvykiui</i> arba <i>Būsenai</i>	Meiliai, gražiai	Prieveiksmis/ būdo apl.
Intensifier	Intensyvumas	Nusako didesnį ar mažesnį <i>Savybės</i> , <i>Būdo</i> arba kito <i>Intensyvumo</i> laipsnį	Labai, truputi, vos	Būdvardis/ Pažyminys
Purpose	Tikslas	<i>Būdo</i> ar <i>Įvykio</i> priežastis	Sotumas, turtingumas	Tikslo aplinkybė

LKS interpretavimą, t. y. nustato, kuri LKS dalis, keičiant standartines sujungimo taisykles, turi būti interpretuojama ir kaip interpretuojama. Šie parametrai ir sudaro pagrindinį konkrečiai kalbai būdingą žinių šaltinį. Pagal nutylėjimą standartinės LKS taisyklės yra tokios kaip ir anglų kalboje, tačiau LKS galima pritaikyti konkrečiai kalbai, kuri turi savas sintaksės bei gramatikos taisykles. LKS pagrindiniai parametrai:

- **Žvaigždutė (\*)** – nurodo, ar LKS dalis turi būti realizuojama sintaksiškai, ar ne. Jei \* yra, reikia, o jei nėra – ne.

*Anglų k. LKS:* enter=[State GO(Thing \*X,Path TO(Position IN(Thing X, Thing \*Z)))] ; I entered the house.

*Lietuvių k. LKS:* Įeiti=[State GO(Thing \*X,Path TO(Position \*IN(Thing X, Thing \*Z)))] ; Aš įėjau į namą.

- **:INT ir :EXT** – pakeičia standartinius loginių subjektų ir argumentų interpretavimo metodus. :INT priverčia taip pažymėtą objektą būti veiksmažodinės frazės (VP) viduje, o :EXT – išorėje.
- **:CAT** – pakeičia sintaksinę objekto kategoriją (vietoj reikšmės pagal nutylėjimą). Pavyzdžiui, *Event* ar *State* tipai pagal nutylėjimą LKS atvaizduoja į veiksmažodžius, o *Thing* į

daiktavardžius. Tačiau konkrečiose kalbose kalbos/sakinio dalį dažnai reikia keisti. Todėl numatyta galimybė įgyti kitą formą pagal konkrečios kalbos taisykles. *Anglų k. LKS*: BE=[Event BE<sub>Possesional</sub>(Thing \*X, TO<sub>Property</sub>(Thing X, Property \*Y))]; – I was told.

*Lietuvių k. LKS*: BE=[Event BE<sub>Possesional</sub>(Thing \*X:EXT, TO<sub>Property</sub>(Thing X, Property \*Y:CAT(VERB):INT))];

Man buvo pasakyta.

- **:DOMOTE ir :PROMOTE** – sakinio temos keitimas, norint pabrėžti arba sumažinti tam tikro LKS leksikinio vieneto tipo įtaką sakinyje. Pvz., Vaikai rinko uogas *miške./* Miške vaikai rinko *uogas*.
- **:CONFLATED** – reikšminiai komponentai yra įtraukiami į LKS, tačiau jų nereikia interpretuoti sintaksiškai.

Šie parametrai leidžia išvengti vertimo nukrypimų. Jie saugo nuo tematinių (gramatinių objektų ir subjektų vietos keitimas skirtingose kalbose), struktūrinių (įvairių prielinksnių bei prieveiksnių įterpimas arba išmetimas) ir kategorinių (kalbos dalies pasikeitimas kitoje kalboje) bei kitokių sintaksinių nukrypimų.

## 2. lentelė Baziniai LKS elementai

<i>Pavadinimas</i>	<i>Apibūdinimas</i>	<i>Pavyzdžiai</i>
<i>Locational</i>	Nurodantis vietą	Petras guli lignoninėje. Jie nuvažiavo į parduotuvę.
<i>Possesional</i>	Teigiantieji	Aš turiu butą. Jis gavo 100 litų.
Identificational	Identifikuojantys	Ji tapo motina. Antanas išliko romus.
Temporal	Laiko	Susitikimas įvyko 13 valandą. Paskaita truko nuo 18 val. iki vėlyvo vakaro.
Circumstantial	Aplinkybės	Dainius ir toliau groja gitara. Remigijus neberašo eilėraščių.
Existential	Egzistavimo	Jonas sukūrė eilėraščių.
Perceptonal	Suvokimo/supratimo	Jie pamatė vaivorykštę. Jis pastebėjo dūmus.
Intentional	Ketinimo	Aš gėriau, nes buvau ištroškęs. Įrašiau muzikinį diskelį.
Instrumental	Instrumentinis	Tomas važiavo su žmona. Martynas apšvietė kambarį su žibintu.

Generavimui iš LKS į lietuvių kalbą naudojami apjungimo metodai, kurie susieja LKS komponentus su atitinkamais sintaksiniais dublikatais. Kadangi LKS yra rekursyvi struktūra, todėl gali būti naudojamas nedaug pakeistas tiesinių struktūrų generavimo algoritmas SHDG (*Semantic Head - Driven Generation*). Būtinai papildymas susijęs su skirtumų sąrašu, t. y. parametrizavimu, naudojamu atvaizduoti išplėstą LKS. SHDG įgyvendina tris pagrindines generatoriaus užduotis – leksikalizaciją (vertimo žodyno naudojimas), struktūrizuotų posakių generavimą ir lingvistinę realizaciją. Ypač patogus gali būti generatorius, kuris efektyviai naudoja gramatiką iš kito galo – pradedant reikšmės atvaizdavimu ir baigiant gramatinių žodžių konstravimu, atitinkančių verčiamo kontekstinę prasmę. Kadangi *interlingua* atveju reikšmė atvaizduojama per LKS formą, todėl generatoriui belieka paimti

verčiamo teksto LKS formos atitikmenį kalboje, į kurią verčiame, ir, naudojant vertimui orientuoto žodyną, sugeneruoti kitos kalbos sakinį.

Verčiant iš anglų į lietuvių kalbą ir atvirkščiai, reikia suformuoti po vieną LKS kiekvienai kalbai, bei parašyti po du algoritmus kiekvienai kalbai: sakinio vertimui iš šaltinio kalbos į LKS ir generavimui iš LKS į tikslo kalbą. Naudojant pažodinį ar transformacinį vertimą, reikalingi tik du algoritmai. Tačiau augant papildomų kalbų skaičiui, *interlingua* vertimo principas reikalauja mažesnio algoritmų skaičiaus (Misevičius, 2002).

Lietuvių kalboje nėra griežtos sakinio tvarkos, nes tą pačią mintį galima pasakyti keliais būdais. LKS būdinga griežta sakinio struktūra vertimui į lietuvių kalbą netrukdo. Nepaisant minėtų *interlingua* privalumų, jos programinis realizavimas yra sudėtingas ir pakankamai painus dėl kelių priežasčių:

- sudėtinga LKS specifikacija, reikalaujanti lingvistinės analizės;
- generavimo problemos iš LKS į tikslo kalbą;
- netaisyklingų sakinių vertimas (geriau tinka pažodinis vertimas);
- painus frazeologizmų vertimas į LKS ir iš jos į tikslo kalbą.

*Interlingua* kalbos panaudojimas būtų paprastesnis tuo atveju, jei lietuvių kalbą būtų galima įjungti į jau esamas vertimo sistemas, naudojančias LKS. Tokiu būdu ji taptų lygiavertė su visomis kitomis vertimo sistemos kalbomis, nes būtų galima versti ne tik iš visų, tos sistemos palaikomų, kalbų, bet ir atvirkščiai. Deja, KV sistemos, paremtos *interlingua* vertimo principu, nepasižymi pakankama vertimo kokybe. Kitas *interlingua* panaudojimo atvejis gali būti intelektualaus teksto skirstymo algoritmo kūrimas, kuris jau pradinėje anglų kalbos teksto analizės fazėje leistų gauti gilesnes kalbos žinias ir leistų sumažinti semantinį daugiaprasmiškumą. Apibendrinus pasakytas mintis, peršasi išvada, kad norint greičiau turėti bent patenkinamos kokybės vertimo sistemą, geriausias būdas naudoti kombinuotą tiesioginį-transformacijos metodą, t. y. sudaryti kiekvieno sakinio gramatinį medį ir jį konvertuoti į atitinkamą lietuvių kalbos medį.

Kompiuterizuoto vertimo sistemų gramatinės analizės komponentai, manipuliuodami gramatinėmis taisyklėmis (Trujillo, 1999; Arnold, 1994; Blekhman, 1998), turi sukurti tikslinėje kalboje sintaksiškai taisyklingą ir pakankamai prasmingą sakinį. Gausiausias taisyklių rinkinys taikomas sakinio dalių išskyrimui ir analizei (veiksniui, tariniui, papildiniui, aplinkybėms ir kt.). Sakinio dalis gali sudaryti vienas ar daugiau žodžių, kurie įvardijami kaip žodžių junginiai. *Sintaksinės sakinio grupės* yra tokie žodžių junginiai, kurie sukuria savarankišką sintaksiškai taisyklingą prasmę sakinyje. Jos nėra vienareikšmis atitikmuo paprastoms sakinio dalims, nors sudaromos pagal tas pačias taisykles. Kartu jos atspindi ir išreiškia tam tikrus prasminius ryšius sakinyje ir palengvina teksto

analizę bei daro paprastesnį ir aiškesnį programinių modulių kūrimą. Verčiant tekstą iš anglų į lietuvių kalbą, sintaksinės sakinio grupės sudaromos pagal anglų, t. y. šaltinio kalbos sakinį. Prieš sudarant sintaksines grupes, sintaksinės analizės komponentas iš kitų kompiuterinio vertimo sistemos žinių šaltinių privalo gauti suskaidytą į sakinius šaltinio kalbos tekstą, kurie, savo ruožtu, turi būti išskaidyti į “žodžio” tipo elementus su atitinkamais atributais.

Šią informaciją turi pateikti žinių šaltinis, kuris atlieka paiešką žodyne. Ketvirtojoje lentelėje pavaizduota vertimo proceso teksto transformavimo informacinė struktūra. Gramatinės analizės modulis, suskaidytą sakinį į “žodžio” tipo vienetus, pagal gautą iš žodyno informaciją ir nustatęs žodžio kalbos dalis, gali kurti sintaksines grupes, jas analizuoti, taikydamas atitinkamas gramatines taisykles. Kadangi įvairiose kalbose kalbos dalių skaičius skiriasi, sintaksinių grupių tipų skaičius taip pat gali skirtis. Į sintaksines grupes žodžiai įtraukiami ir jungiami pagal prasmę ir vaidmenį sakinyje, todėl ne visoms kalbos dalims būtina sudaryti tokias grupes. Pavyzdžiui, lietuvių kalboje netgirieveiksmis turi ryšių su kitais sakinio žodžiais:

$$\begin{array}{ccc} daug & \xrightarrow{ko?} & lai\check{s} \text{ k} \\ \text{prieveiksmis} & & \text{daiktav.} \end{array}$$

Todėlrieveiksmiui lietuvių kalboje tenka sudaryti atskirą sintaksinę grupę. Pagal tai, kuri kalbos dalis yra pagrindinis sintaksinės grupės dėmuo, skiriamos *daiktavardžio*, *veiksmažodžio*, *prielinksnio*, *prieveiksmio* sintaksinės grupės (Dobrovolskis, 2000; Ambrazas, 1997).

Trumpai charakterizuosime pagrindines sintaksines sakinio grupes (anglų–lietuvių kalbų vertimo atveju).

- ✓ Daiktavardžio grupė (Noun phrase) – daiktavardis ir įeinančios į grupę kalbos dalys: būdvardis, determinantai, artikeliai, įvardis, skaitvardis. *Pastaba:* jei sakinyje iš eilės eina dvi daiktavardžio grupės atskirtos jungtuku ar prieveiksmiu, jos gali būti sujungtos į vieną, įtraukiant ir jungiamąjį žodį.
- ✓ Veiksmažodžio grupė (Verb phrase) – veiksmažodis ir įeinančios į grupę kalbos dalys: papildantis veiksmažodis anglų kalboje (pvz., “could”, “may”). *Pastaba:* jei sakinyje iš eilės eina dvi veiksmažodžio grupės, atskirtos jungtuku, jos gali būti sujungtos į vieną, įtraukiant ir jungiamąjį žodį.
- ✓ Prielinksnio grupė (Preposition phrase) – prielinksnis. *Pastaba:* prielinksnio grupė seka prieš daiktavardžio grupę.
- ✓ Prieveiksmio grupė (Adverb phrase) – prieveiksmis. Jų grupės gali būti sujungtos į vieną, jei jos atskirtos jungtuku.
- ✓ X grupė skirta jungti kitų kalbos dalių žodžius ar frazes, kurie nepatenka į kitas grupes. Tai gali būti pavadinimai, ženklai ar nekaitomos frazės sakinyje, taip pat nepažymėti kita kalbos dalimi žodžiai ankstesniame lygmenyje.

**3. lentelė Žodžio atributai**

<i>Atributas</i>	<i>Paaiškinimas</i>
Žodis šaltinio kalboje (anglų)	Žodis suskaidytame šaltinio sakinyje
Vertimas tikslinėje kalboje (lietuvių)	Šaltinio kalbos žodžio vertimas gautas iš žodyno
Žodžio kamienas tikslinėje kalboje	Žodžio kamienas tikslinėje kalboje, be galūnės
Kalbos dalis	Sintaksinių grupių sudarymui reikia žinoti: daiktavardį, įvardį, veiksmažodį, skaitvardį,rieveiksmį, būdvardį, determinantą, prielinksnį, jungtuką, skyrikį. Kitos kalbos dalys gali būti priskirtos neįvardytoms kalbos dalims, jei jos ypatingos reikšmės sakinio analizei neturi.
Linksnis	Daiktavardis ir būdvardis lietuvių kalboje kaitomas linksniais.
Skaičius	Skaičius – būdvardžiui ar daiktavardžiui (keičiant žodžio formą)
Giminė	Giminė – būdvardžiui ar daiktavardžiui (keičiant žodžio formą)
Laikas	Lietuvių ir anglų kalboje veiksmažodis laikui žymėti turi skirtingas sudarymo taisykles
Asmuo	Veiksmažodis asmenuojamas.
Nuoroda į žodyną	Nuoroda į žodyną, spartesnei žodžio paieškai.

Visi sakinio žodžiai turi būti priskirti kokiam nors grupei. Sintaksinių grupių sudarymo pavyzdys, verčiant iš anglų į lietuvių kalbą, pateiktas penktoje lentelėje. Daiktavardžio grupės gali būti dviejų tipų *objektas* arba *subjektas*. Grupė *subjektas* atlieka veiksnio vaidmenį sakinyje, grupė *objektas* – papildinio, pažyminio ar aplinkybių vaidmenį.

Sudarytoms sintaksinėms sakinio grupėms yra paprasčiau taikyti gramatines, morfologines ir sakinio skyrybos taisykles. Pvz., pagal sakinio darybos taisyklę sintaksiškai teisingą sakinį būtinai sudaro daiktavardžio ir veiksmažodžio grupės, kitos grupės yra papildomos. Šiai taisyklei yra nesvarbu kiek sakinyje yra žodžių ir kokia jų tvarka, tačiau ji sako, jog sakinyje būtinai turi būti bent vienas daiktavardis, įeinantis į daiktavardžio grupę, ir bent vienas veiksmažodis, t. y.

#### 4. lentelė Sintaksinių grupių sudarymo ir taisyklių taikymo joms pavyzdys

Šaltinio sakinys:	Little boys went to a big school.							
Informacija iš žodyno	Little	Boys	Went	to	a	big	school	.
	būdvardis	daiktavardis	veiksmažodis	prielinksnis	det	būdvardis	daiktavardis	Skyr
Grupių sudarymas	Mažas	Berniukai	Ėjo	Į		didelis	mokykla	
	daiktavardžio grupė		veiksmažodžio grupė	prielinksnio grupė	daiktavardžio grupė			x gr.
Taisyklių grupėms taikymas	Little	Boys	Went	to	a	big	school	.
	būdvardis	daiktavardis	veiksmažodis	prielinksnis	det	būdvardis	daiktavardis	skyr
	daiktavardžio gr. subjektas		veiksm. gr.	priel. gr.	daiktavardžio. gr. objektas			
Tikslo sakinys	Maži	Berniukai	Ėjo	Į		didelę	mokyklą	.
	Maži berniukai ėjo į didelę mokyklą.							

S→ D.Gr.+V.Gr. arba S→D.Gr+V.Gr+[Priel. Gr.].

Šioms grupėms patogiu taikyti sintaksinio derinimo, valdymo, šliejimo ryšių taisykles:

- derinimas – grupės priklausomojo žodžio giminės, skaičiaus ir linksnio derinimas pagal pagrindinį žodį. Pvz., daiktavardžio grupėje linksniai, giminė derinami pagal pagrindinį žodį – daiktavardį: {maži berniukai} d.gr., {didelę mokyklą};
- valdymas – pagrindinis žodis reikalauja tam tikro priklausomųjų žodžio linksnio. Pavyzdžiui:

$$\underset{\text{veiksm. gr.}}{\text{raš o}} \xrightarrow{\text{kq}} \underset{\text{daiktav. gr.}}{\text{laiš k}}, \quad \underset{\text{priel. gr.}}{\text{Į}} \xrightarrow{\text{kq}} \underset{\text{daiktav. grupė}}{\text{didelę mokyklą}},$$

- šliejimas, kai su pagrindiniu žodžiu pagal prasmę susiejamas nekaitomas žodis. Pavyzdžiui:

$$\underset{\text{prieveiksm. gr.}}{\text{daug}} \xrightarrow{\text{ko?}} \underset{\text{daiktav. gr.}}{\text{mandagių laišk}}$$

Sintaksinėms grupėms taikyti taisykles, kurios susiję su daiktavardžių grupėmis, subjektais arba objektais yra paprasčiau, kai subjektas gali valdyti objekto linksnį, giminę ar skaičių:

Jie {daiktav. gr. subjektas} yra {veiksm. gr} gražūs berniukai {daiktav. gr. objektas}

## 2.6 Kompiuterizuoto vertimo proceso funkcinė struktūra

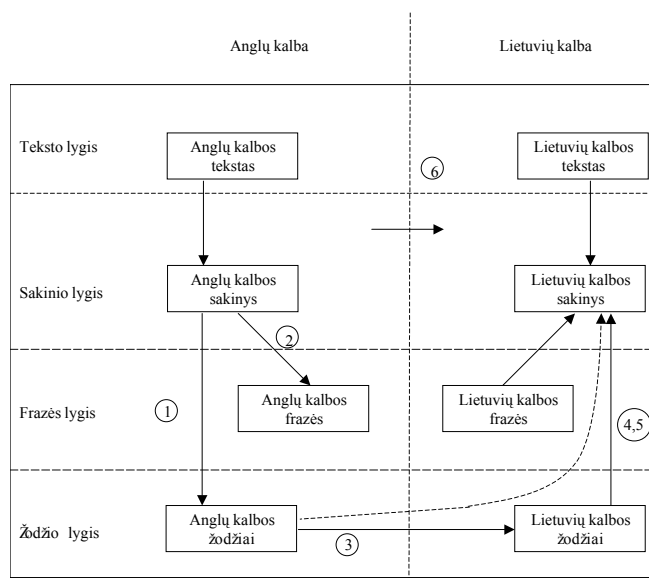
Vertimo proceso eigą galima aiškiau suvokti, panagrinėjus elementaraus anglų kalbos sakinio (išeities teksto) “A book is on the table.” vertimą į lietuvių kalbą (žr. 5 lent.). Pirmojoje vertimo fazėje originalo (išeities) anglų kalbos sakinyje skaidomas į atskirus žodžius. Pažodžiui išvertus visus žodžius,

nustatoma jų prasmė, forma, kalbos dalis ir pan. Kituose žingsniuose, identifikavus žodžių funkciją sakinyje, formuojamas semantiškai adekvatus originalui sakinyje bei gramatiškai derinami žodžiai.

**5. lentelė Anglų kalbos sakinio vertimo fazės**

<i>Sakinys originalo klb.</i>	<i>A</i>	<i>Book</i>	<i>Is</i>	<i>on</i>	<i>the</i>	<i>table</i>
Reikšmė		<i>Knyga</i>	<i>yra (būti)</i>	<i>ant</i>		<i>stalas</i>
Kalbos dalis	žym. Artikelis	Daiktavardis	veiksmažodis	prielinksnis	žym. artikelis	daiktavardis
Forma			esam. l., vnsk., 3 asm.			viet. linksnis
Funkcija sakinyje		Veiksny	Tarinys			papildinys
Sakinys tikslinėje klb.		<i>Knyga</i>	<i>Yra</i>	<i>ant</i>		<i>stalo</i>

Šis paprastas pavyzdys, nors ir neatskleidžia vertimo proceso sudėtingumo, tačiau pakankamai vaizdžiai iliustruoja pagrindinius vertimo principus bei etapus. Pirmajame paveikslėlyje vaizduojami apibendrinti pagrindiniai vertimo žingsniai ir atitinkama funkcinė schema.

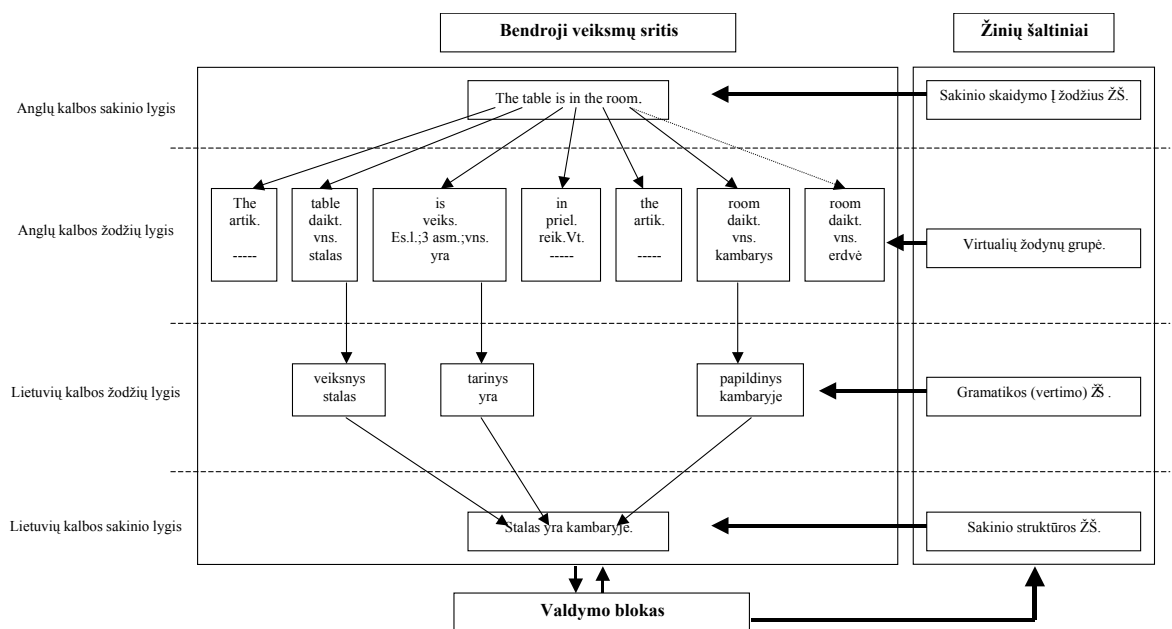


**7 pav. Vertimo proceso funkcinė schema**

Funkcinėje schemoje skaičiais parodyta vertimo proceso pagrindinės fazės arba etapai:

- 1 – pradinis sakinio paruošimas (skirstymas, analizė, žodžių atributų išskyrimas);
- 2 – frazių (frazėologizmų) paieška, analizė, vertimas;
- 3 – pažodinė analizė (kalbos dalies nustatymas, originalo kalbos žodžio formos nustatymas, pažodinis vertimas);
- 4 – sakinio vertimas (žodžių atitiktens ir funkcionalumo sakiniuose nustatymas, jų suderinimas);
- 5 – galutinis tikslinės kalbos sakinio sutvarkymas (žodžių tvarka, skyrybos ženklai);
- 6 – semantinių ir pragmatinių žinių gavyba iš tekstynų ir tekstų suderinamumo „filtrų“ panaudojimas.

Vertimo proceso analizė leidžia skirti kai kurias būdingas jo ypatybes. Tai sudėtingas uždavinys: gausu žodžių ir frazių vertimo variantų, prasmė priklauso nuo konteksto, reikalingos papildomos probleminės žinios ir pan. Nesunku pastebėti, kad vertimas iš vienos kalbos į kitą turi aiškia hierarchiją, t. y. vertimo objektai (sakiniai, frazės, žodžiai) yra išsidėstę skirtinguose hierarchiniuose duomenų pateikimo lygmenyse. Siekiant paspartinti vertimo procesą, efektyviau išnaudoti kompiuterio resursus, dalis kompiuterizuoto vertimo žingsnių gali būti atliekami lygiagrečiai (pvz., žodžių ir frazių apdorojimas), kiekvienas vertimo etapas arba fazė gali būti nagrinėjami kaip atskiri uždaviniai, o juos vykduojantieji programiniai moduliai gali būti aprašomi ir realizuojami nepriklausomai vienas nuo kito. Įvertinus išvardytus ypatumus, kompiuterinio vertimo konceptualiajam modeliui siūloma taikyti hierarchinę architektūrą, kuri žinių inžinerijoje atitinka, vadinamąjį sprendimų lentos.



8 pav. Kompiuterizuotos vertimo sistemos konceptualusis modelis



## 2.7 Gramatinis šaltinio teksto nagrinėjimas.

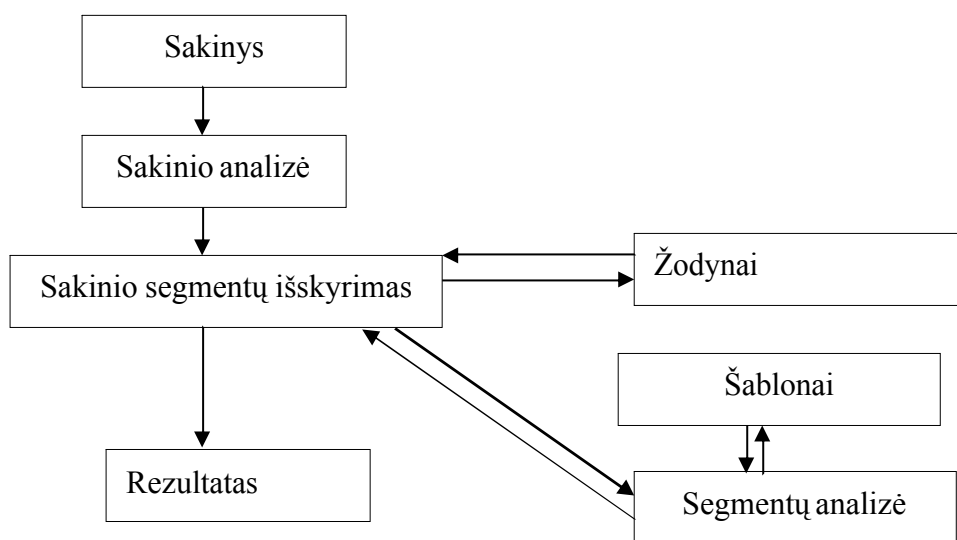
### 2.7.1 Segmentinis sakinio analizės būdas

Nustatoma kiekvieno žodžio sakinio dalis: daiktavardis, veiksmažodis, būdvardis,rieveksmis, dalelytė, Įvardis, skaitmuo, prielinksnis, jungtukas ar jaustukas.

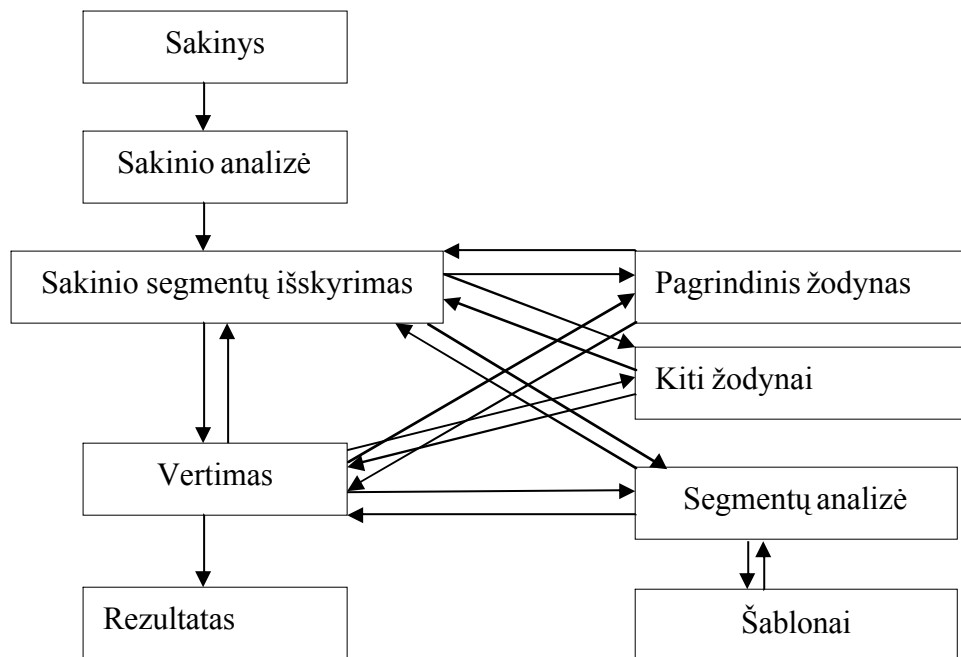
Pagal tam tikras taisykles sakiny s skirstomas į segmentus. Į segmentus skirstoma pagal šablonus imama dažniausiai 2-3-4 ar ilgesnės žodžių grupės atitinkančios šablona, tada ta žodžių grupė pažymima kaip atpažinta sakinio dalis ir kuris šablonas jei priklauso. Taip visas sakiny s suskirstomas šablonais (9 pav.) Neretai pasitaiko, kad į segmentą įeina tik vienas žodis.

Išanalizavus visą sakinį vykdomas jo vertimas. Geresnės vertimo kokybės pasiekama (10 pav.) jei KV blokas atsižvelgia į sakinių analizės bloko atmestus šablonus.

Pvz.: On top of it. Šis pavyzdys kompiuteriui gana sudėtingas. Čia galima išskirti 3 segmentus „on top“, „top of“, „of it“, tačiau kiekvienas iš jų turi po du žodžius t.y. gali būti „on“ „top“ „of“ „it“. Šioms grupėms priklauso tokie junginiai „on top“, „on daiktavardis“, „on you“, „top secret“, „top most“, „of if“, „of all“, „it is“, „it takes“ (čia tik keli pvz.). Taigi kompiuteris gali išskirti 3 segmentus po 2 žodžius. Teisingai išversti padės vertimo algoritmas kuris įvertins „on top“ ir „of it“ segmentus palankiau, negu „on“, „top of“ ir „it“.



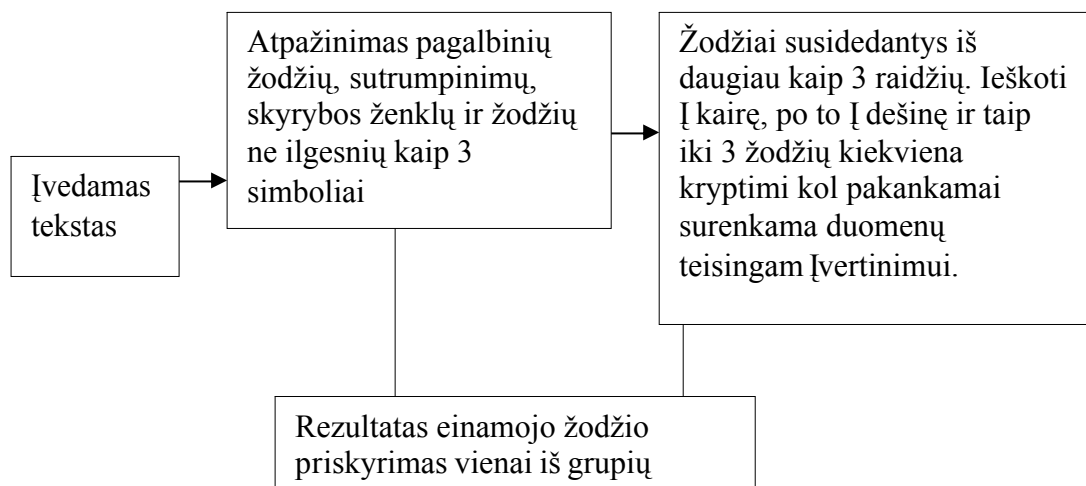
9 pav. Segmentinė sakinio analizės schema



10 pav. KV schema naudojanti segmentinę sakinio analizę

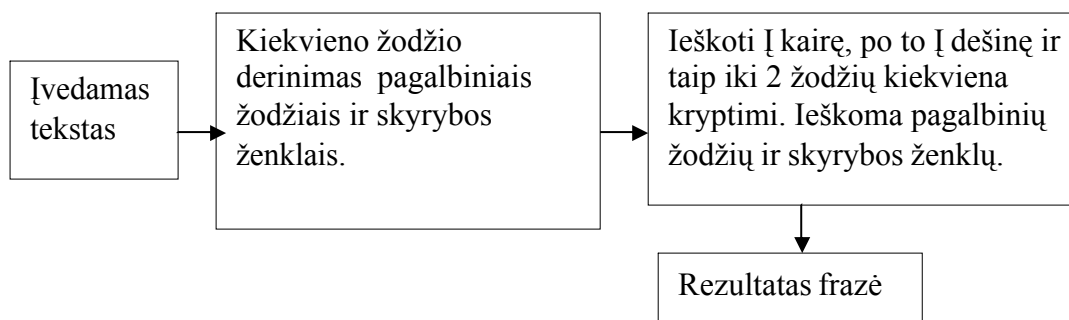
## 2.7.2 Automatinis sakinio analizės būdas

Automatinio vertimo programos pagrindinė idėja yra žodžio informacijos rinkimas pagal jo buvimo vietą tame sakinyje. Tai labai naudinga, kai žodyne yra keletas ar net dešimtys įvairiausių vertimo variantų. Automatinis sakinio analizės būdas labai pravartus, kai žodyne mes neturime informacijos apie vieną ar kitą žodį sakinyje, todėl jis yra efektingiausias tokiu atveju (11 pav.).



11 pav. Žodžio sakinyje informacijos rinkimas

Sakinyje gali būti išskiriamos frazės pagal tam tikras lenteles.



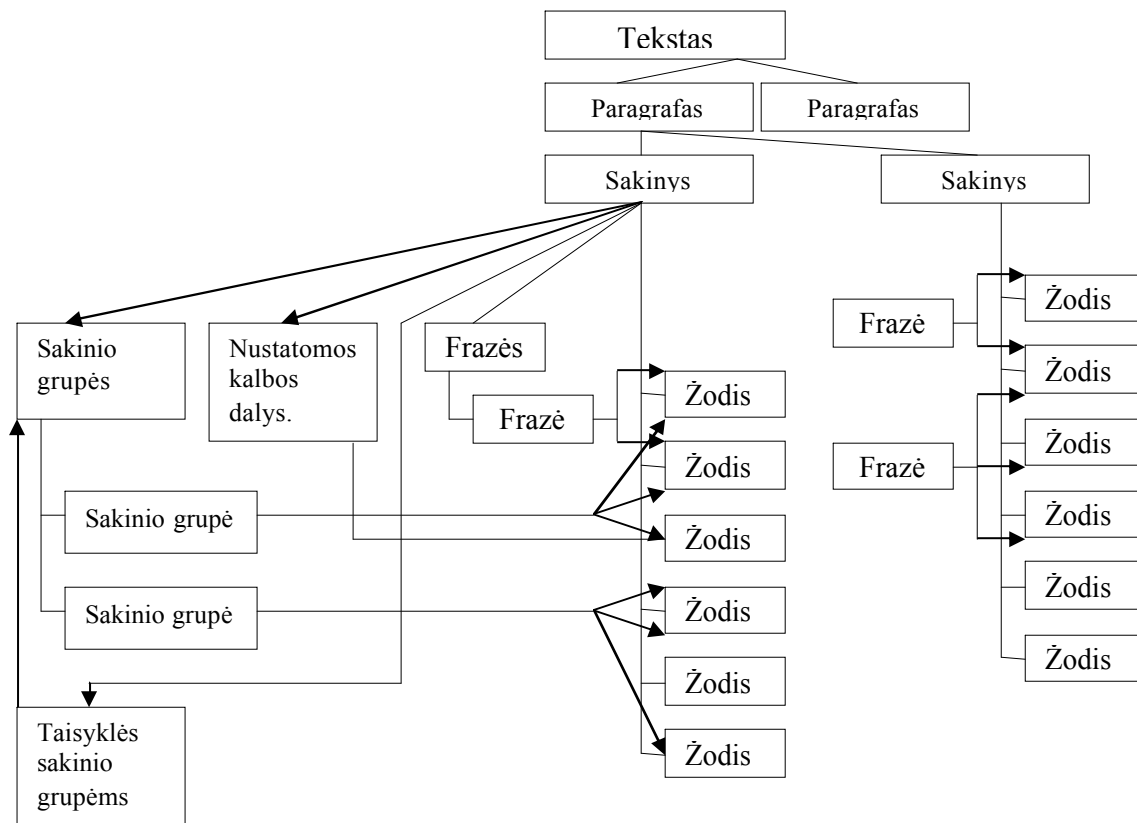
12 pav. Frazijų išskyrimo algoritmas

Šis analizės būdas yra itin sudėtingas. Jis duoda gana gerus rezultatus kai „šablonais skaidoma į segmentus“ algoritmas neranda tinkamų šablonų, tačiau daugeliu atveju jo rezultatas nebus toks tikslus kaip segmentiniu sakinio analizės būdu.

Čia buvo pavaizduoti tik kelios paprastos schemas. Komeraciniai produktai naudoja daug sudėtingesnius algoritmus, kurių kiekvienas turi didelės apimties būsenų, išimčių lenteles. Daugelis iš šių algoritmų turi ir išimčių, kurias reikia numatyti, kitaip gero vertimo rezultato galime ir nepasiekti.

### 2.7.3 Taisyklių nustatymo sakinio analizės būdas

Pateiktas analizės būdas, kai tekstas skaidomas paragrafais, paragrafai sakiniiais o sakiniai žodžiais. Kiekvienas iš šių sakinių gali turėti frazes. Frazijų gali būti daug, kiekvienai iš jų priskiriami žodžiai. Skirtingoms frazėms priskirti žodžiai gali dubliuotis. Toliau kiekvienam žodžiui nustatoma kalbos dalis, pagal išskirtas kalbos dalis, bei sakinio struktūrą, sakinyje sudaromos sakinio grupės, o po to nustatomos ir vertimo taisyklės sakinio grupėms.

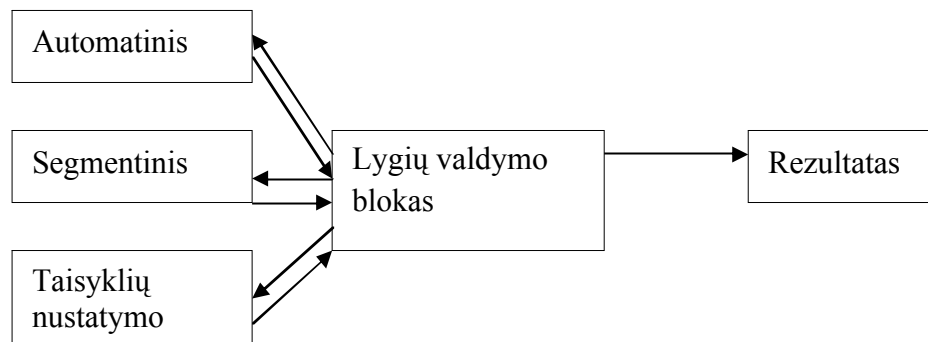


13 pav. Teksto analizė

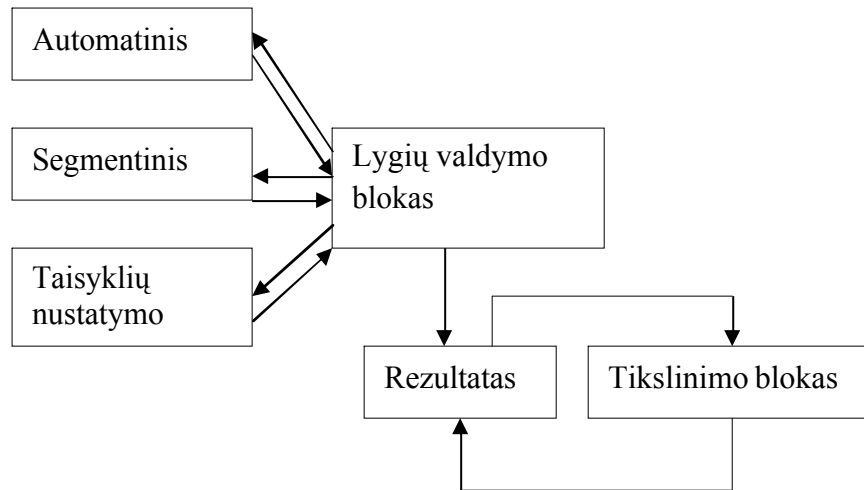
Pavyzdys, kaip vyksta vertimas po „taisyklių nustatymo“ sakinio analizės būdo.

#### 2.7.4 Hibridinis sakinio analizės būdas

Hibridinis – toks sakinio analizės būdas, kuris naudoja algoritmus automatinio, segmentinio ir taisyklių ir kitus nustatymo algoritmus, sulieja dviejų ar daugiau algoritmų pateiktą informaciją į vieną. Hibridiniai sakinio analizės metodai skirstomi į dvi grupes: euristinius ir iteracinius 14 ir 15 paveikslai.



14 pav. Euristinė hibridinio analizės metodų valdymo schema



**15 pav. Iteracinė hibridinio analizės metodų valdymo schema**

Euristinis – toks sakinio analizės būdas, kai lygių valdymo blokas, pagal jam užduotas taisykles, iš karto parenka labiausiai tinkamą schemą kiekvienam lygiui.

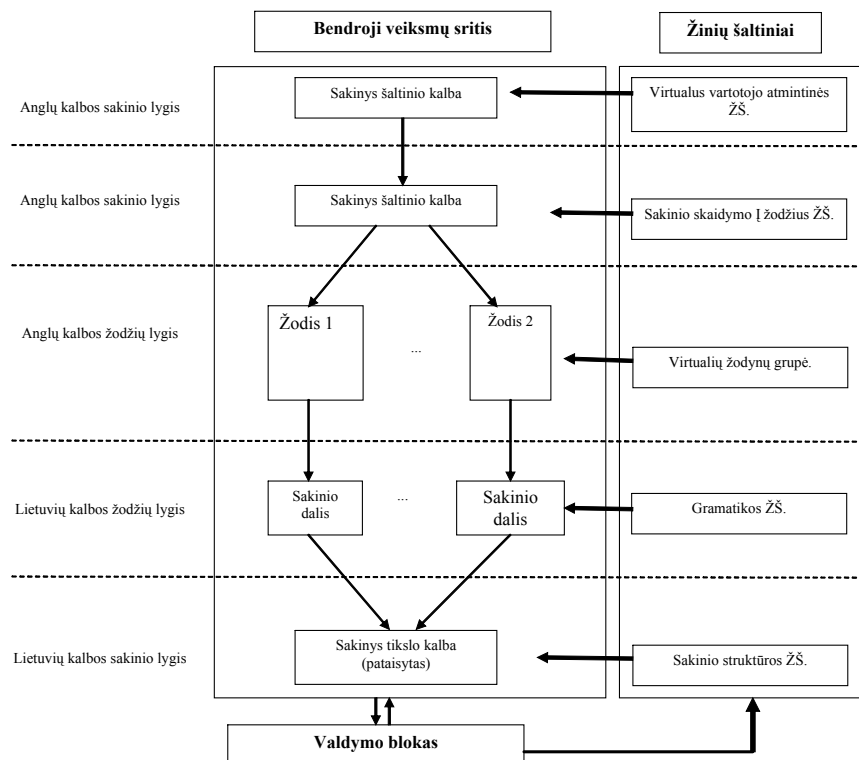
Euristinis nuo iteracinio skiriasi tuo, kad iteracinis algoritmas daro rezultato tikslinimus tol kol pasiekiamas toks lygis, kai prieštarauja kuo mažesniai kiekiui taisyklių su mažiausiais svoriais

### 3 KV sistema iš anglų į lietuvių kalbą

Šiame skyriuje pateikiamas sprendimo modelis. Aiškinama kaip jis veikia. Pateikiamas sistemos veikimas, jos vaizdas. Pateikiami sistemos panaudotų ŽŠ veikimo principai. Vertimo galimybės ir pavyzdžiai.

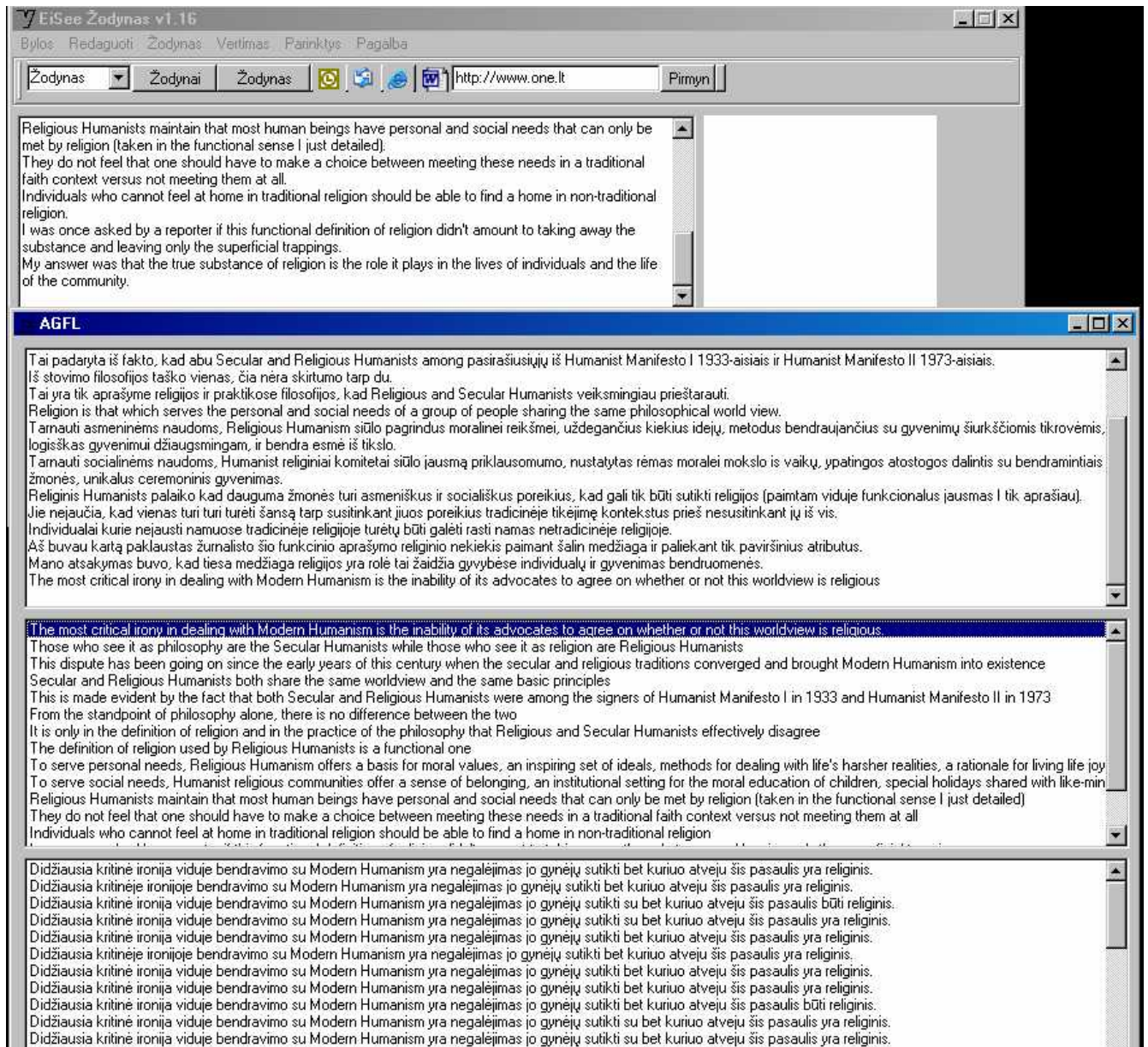
#### 3.1 KV sistemos modelis

Kuriant sistemą buvo pasinaudota sprendimų lentos modeliu. 16 pav. pateiktas demonstruojamas sprendimų lentos modelis. Pirmame lygyje pateiktas vartotojo atmintinės žinių šaltinis, jo pagalba sakinio vertimas gali būti iš karto randamas apeinant kitus lygius, tačiau jei žinių šaltinis savo duomenų bazėje neranda, tuomet sakinys yra skaidomas į žodžius „sakinio skaidymo į žodžius ŽŠ“ ir ieškoma kiekvieno žodžio vertimo, neradus prašoma vartotojo įvesti. Anglų kalbos žodžių lygyje mes turime suskaidytus angliškus žodžius. Jei vertimas vyksta pažodiniu režimu tai gramatikos nagrinėjimas yra praleidžiamas. Nurodžius naudoti sakinio analizę, atliekama analizė AGFL sistemos pagalba. Lietuvių kalbos sakinio lygyje pasinaudojus „sakinio struktūros ŽŠ“ sugeneruojamas sakinys. Kadangi sakinio analizė yra tikimybinė, o kartais žodyne būna daug vieno žodžio vertimo variantų, tai sugeneruojami keli ar daugiau sakinių ir pateikiami vartotojui pasirinkti.



16 pav. Programos ekrano vaizdas verčiant su sakinio analize

Programos ekrano vaizdas. Aukščiausiam lange yra angliškas tekstas. Žemiau pakeiktas to teksto vertimas. Trečiame nuo viršaus lange yra sakiniai kuriuos vertė sistema. Pasirinkus bet kurį sakinį, vartotojui pateikiamas žemiausiame lange to sakinio galimi vertimo variantai. Variantai pateikiami, tam, kad vartotojas galėtų išsirinkti taisyklingiausią sakinį iš tų, kuriuos programa gali pateikti. Programos ekrano vaizdas.



17 pav. Programos ekrano vaizdas verčiant su sakinių analize

Naujai įdiegta sistema su sakinių sintaksine ir semantine analize pakėlė vertimo kokybę. Programa su vartotojo pagalba gali išversti apytiksliai kas ketvirtą sakinį, kad programos vartotojas galėtų jį suprasti, nors sakinyvis nebūtinai yra visiškai taisyklingas. Programą galima rasti adresu:

<http://fortas.ktu.lt/~sipavi/mokslai/>.

## 3.2 Žodynų – leksikonų pildymas.

Labai svarbu, kad žodynus galima būtų pildyti. Tokie žodynai vadinami virtualiais. Žodynų pildymas yra tikslingas, nes kasdien mes sutinkame naujų ir vertimo programai nežinomų žodžių (jų darinių, frazių). Žodynus tikslinga pildyti pagal tematikas pvz.: kompiuterinius, medicininius, techninius ir kitus žodynus. Žodynų klasifikavimas ir teisingas jų parinkimas vertimo metu, neleidžia medikui verčiančiam straipsnį iš medicinos žurnalo, rasti žodį iš mechanikos srities ir pan.

Lietuvių kalbos dalys (būdvardis, veiksmažodis, daiktavardis ir kt.) turi labai daug atributų ir juos visus reikia užpildyti (18-22 pav.), jei norime, kad programa teisingai juos suprastų ir įvertintų vertimo metu. Atributų pildymas ilgai užtrunka, todėl vartotojui reikia padėti atlikti tai per trumpiausią laiką, automatizuojant tam tikrus laukus (pvz.: automatinis daugiskaitos generavimas, linkšnių generavimas pasirinkus linkniuotes) pateiktas daiktavardžio, veiksmažodžio, būdvardžio, įvardžio ir skaitvardžio atributų pildymo pavyzdžiai.

Detali informacija

Daiktavardis | Būdvardis | Veiksmažodis | Ivardis | Skaitvardis | Kita kalbos dalis

Liefuviu1 ::Anglu1 ::Lik

Angliškas daiktavardis

Vienaskaita  
hous

Daugiskaita  
houses

Skaičiavimas  
 Abu skaičiai  
 Tik vienaskaita  
 Tik daugiskaita

Tikrinis  
 Netikrinis  
 Tikrinis

Lietuviškas daiktavardis

Žodžio kamienas be galūnės  
nam

Ką nusako šis daiktavardis (kategorija)  
Gyva būtybė (žmogus)

Giminė  
 Vyriška  
 Moteriška  
 Abiejų giminė?

Skaičiavimas  
 Abu skaičiai  
 Tik vienaskaita (sukrus)  
 Tik daugiskaita (žirkl?is)

Linksniuotė  
Parinkite kamienui galūnę  
-as, -o, -ams | vyras, vyrams

namas, namo, namui, namą, namu, name

namai, namų, namams, namus, namais, namuose

Idėti | Atšaukti

Idėti | Atnaujinti | Ištrinti

Idėti | Atnaujinti | Ištrinti

18 pav. Daiktavardžio atributų pildymas



19 pav. Būdvardžio atributų pildymas

Pasirinkus kalbos dalį „**Veiksmazodis**“ reikės įvesti angliškojo žodžio I, II, III formas o taip pat ir II asmens esamąjį laiką. Reikia įvesti lietuviško žodžio kamieną, bendrąsias galūnes ir kas nusakoma šiuo veiksmazodžiu. Dar reikės pasirinkti esamojo, būtojo ir būsimojo laikų galūnes.

20 pav. Veiksmazodžio atributų pildymas

Pasirinkus kalbos dalį „Ivardis“ reikės įvesti angliškąjį žodį bei lietuvišką žodį, jam parinkti tipą, giminę ir skaičių.

21 pav. Ivardžio atributų pildymas

Pasirinkus kalbos dalį „Skaitvardis“ reikės įvesti angliškojo žodžio vienaskaita ir daugiskaitą (jei tokia yra), būtina kad įvestumėte ir jo skaitinę reikšmę.

22 pav. Skaitvardžio atributų pildymas

### 3.3 Gramatinis sakinių nagrinėjimas

Sintaksė yra gramatikos dalis, nagrinėjanti žodžių tarpusavio ryšius, sakinių sandarą, jų sudarymo būdus ir taisykles. Sintaksei taip pat rūpi ir tekstas, sakinių sąsajos jame.

Pagrindinės sintaksės sąvokos — žodžių junginys, sakinys (tiesioginis, skatinamasis ir klausiamasis; išplėstinis ir neišplėstinis; vienanaris ir dvinaris; asmeninis ir beasmenis; pilnasis ir nepilnasis; vientisinis ir sudėtinis), sakinio dalys (pagrindinės — veiksnys ir tarinys; antrininkės — papildinys, aplinkybės ir pažyminys), kreipinys, Įterpiniai, tiesioginė ir netiesioginė kalba.

Sintaksė tiria žodžių santykius, jų jungimo būdus, taip pat ir sakinių modelius, sudėtinių sakinių sudarymo būdus, jų rūšis. Pastaruoju metu sintaksei priskiriamas ir tekstas, t. y. sakinių sąsaja.

Sintaksės tyrimo esmę visada sudaro žodžių tarpusavio santykiai, žodžių ar sakinių jungimo taisyklės. Žinoma, sintaksei rūpi ir žodžio formos, bet tik tiek, kiek jos gali atlikti tam tikras sintaksines funkcijas.

**Žodžių tvarka sakinyje.** Iš pirmo žvilgsnio gali atrodyti, kad lietuvių kalbos sakinio žodžių tvarka visiškai laisva. Antai sakinį *Viesulas išrovė medį* galima pasakyti penkeriopai, ir visi variantai bus suprantami:

*Viesulas medį išrovė.*

*Išrovė viesulas medį.*

*Medį viesulas išrovė.*

*Išrovė medį viesulas.*

*Medį iš rovė viesulas.*

Taip yra todėl, kad lietuvių kalbos žodžių ryšį sakinyje rodo galūnės. Tačiau kurį variantą tikslingiausia pasirinkti kalbant arba rašant, priklauso nuo to, kas klausytojui ar skaitytojui yra žinoma iš ankstesnio teksto. Jei, tarkim, iš jo yra aišku, kad buvo kilęs viesulas ir norima pranešti, kad jis išrovė medį, o ne ką kitą, žodžių tvarka būtų tokia: *Viesulas išrovė medį*. Ir, priešingai, jei būtų žinoma, kad medis išrautas, ir rūpėtų pasakyti, kad tai padarė viesulas, reiktų jau tokios žodžių tvarkos: *Medį išrovė viesulas*. Taigi sakinio pradžioje sakome tai, ką žinome iš ankstesnių sakinių, o pabaigoje — ką naujo pranešame.

Kadangi nauja paprastai pasakoma tariniu su jam priklausomomis sakinio dalimis, ši pagrindinė sakinio dalis dažniausiai eina po veiksnio.

*Mikutis šypteli Juozapėliui. Paskutinės ašaros džiūsta jo skruostuose, o akys giedros ir geros. Jie jau nebe priešai, bet draugai.*

Įprastinė papildinio vieta sakinyje — po tarinio (*Mama rašo laišką*). Po tarinio dažnai eina ir daugumas aplinkybių (*Kelias vingiuoja tolyn. Avys laikomos vilnai.*). Būdo aplinkybė, priešingai, paprastai eina prieš tarinį (*Jonas neryžtingai atsistoja.*). Neretai sakinio pradžioje pasakomos vietos ir laiko aplinkybės (*Vakarop lietus nurimo. Miške giedėjo paukščiai.*).

Įsidėmėtina pažyminio vieta sakinyje. Kilmininku išreikštas nederinamasis pažyminys eina prieš pažymimajį žodį:

*Nakties rasą sugėrė įkaitę akmenys.*

*Vėjas čiužena rudens lapus.*

Prieš pažymimajį žodį eina ir derinamasis pažyminys:

*Drėgna žemė kv epėjo pavasariu.*

*Kitą rytą upėje jau plaukė pirmas išas.*

Įnagininku išreikštas nederinamasis pažyminys eina po pažymimojo žodžio:

*Pasirodė vyras šiaudine skrybėle.*

*Būgną mušė žmogus su geltonomis sagomis.*

Kartais Įprastinė žodžių tvarka keičiama norint kurį nors žodį pabrėžti.

*Gražu buvo neapsakomai.* (Plg.: *Buvo neapsakomai gražu.* )

*0 jaunosios dienos mano!* (Plg.: *0 mano jaunosios dienos!* )

*Ten už upių plačių spindi mūsų pulkai.*

*Dūmas tėvynės už svetimą ugnį šviesesnis.*

Toks Įprastinės žodžių tvarkos pakeitimas vadinamas inversija.

Kalbant ir skaitant didesnę prasminę krūvę turinčius žodžius galima išskirti loginiu kirčiu, nekeičiant žodžių tvarkos.

*Viesulas išrovė medį.* (Viesulas, o ne darbininkai ar kas kitas.)

*Viesulas išrovė medį.* (Išrovė, o ne nulaužė ar pan.)

Kadangi anglų kalbos sakinių analizės programų jau yra sukurta, netikslinga kurti iš naujo, naudingiau jau sukurti tinkamą analizę lietuvių kalbai. „Nijmegen“ universiteto sukurta anglų kalbos sakinių analizės sistema (Affix Grammars over a Finite Lattice) naudoju tolesniems tyrimams. AGFL – sistema kaip tik gali būti tinkama tolesniems tyrimams. Ji yra nemokama, atviro kodo, gali redaguoti visi kas sugeba tik ir Įsipareigojama, kad naujausius pataisymus gautų autoriai bei panaudota AGFL sistema, ar jos dalis negali būti naudojama komerciniais tikslais.

AGFL sistema tinkamiausia iš visų žinomų dėl keleto pagrindinių priežasčių:

1. Visų pirma tai kad ji yra atviro kodo, ją galima lengvai redaguoti.

2. Jos kokybė patenkinama. Pati sistema turi informaciją apie daugiau kaip 30000 žodžių, kurią po to naudoja sakinio analizei.
3. Pateiktos sakinio analizės yra tikimybinės. Pateikia ne vien tik pačią tinkamiausią, bet ir mažiau tikėtiną analizę. Tai leidžia įdiegti vartotojo pagalbos priemones geresnės kokybės pasiekimui.

Įvedus toki sakinį “ Little boys go to school”. Gauname tokią analizę.

segment

sentence

statement

simple statement

SVOC phrase

opt circumstances

circumstance

adverb

LEX\_ADV

LADV

ADV

"little"

opt circumstances

subject(plur, third)

NP(plur, third, nom)

noun phrase(plur, third, nom)

noun part(plur, third, nom)

noun group(plur, third, nom)

noun kernel(plur)

LEX\_NOUN(plur)

LNOUN(plur)

NOUN(plur)

"boys"

VOC phrase(plur, third)

verb group(intr, none, to, plur, third)

verb form(intr, none, to, plur, third)

LEX\_VERBI(to, intr)

VERBI(to, intr)

"go"

OC phrase(intr, none, to)

C phrase(to)

PP(to)

opt adverbs

LEX\_PREPOS(to)

PREPOS(to)

"to"

noun phrase(sing, third, dat)

noun part(sing, third, dat)

noun group(sing, third, dat)

noun kernel(sing)

LEX\_NOUN(sing)

LNOUN(sing)

NOUN(sing)

"school"  
opt circumstances

### 3.4 Vertimai iš anglų į lietuvių kalbą

Prototipinėje vertimo iš anglų kalbos į lietuvių kalbą KV programoje buvo realizuotos sintaksinė semantinė sakinio analizė, kuri skirta įvairiems tekstams (pvz., dokumentų) vertimui. Įdiegtas gramatinės analizės komponentas su sintaksinių sakinio grupių taikymu, žymiai pagerino lietuviškų tekstų vertimo kokybę. Šiuo metu gramatinės analizės komponentas jau naudoja per 20 taisyklių įvairioms sintaksinėms grupėms. Programoje įdiegtas universalus virtualaus žodyno modelis, kuris gali būti papildomas atitinkamais atributais priklausomai nuo to kuriai kalbos daliai priklauso. Nepriklausomi išskirtinių žodžių ir morfologinės analizės komponentai reglamentuoja lietuviškų žodžių linksniuotųjų ir asmenuotųjų taikymo tvarką. Įdiegtas frazeologizmų analizės komponentas. Įdiegtas vartotojo atmintinės modulius leidžia apeiti visą vertimo procesą, jei tokį sakinį vartotojas yra vertęs ankstesniuose savo darbuose.

Vertimai pagal savo prasmę ir gramatinę struktūrą, kad būtų adekvačiai įvertinta kokybė reiktų vertinti pagal sekančius kriterijus:

- Sakinys turi prasmę ir yra visiškai teisingas.
- Sakinys turi prasmę, bet sakinio struktūra neteisinga.
- Sakinys turi prasmę, bet turėtų būti panaudoti kiti (sinonimai) žodžiai.
- Prasmę galima suprasti, bet dėl sintaksiškai neteisingo sakinio, teisingai suprasti sunku.
- Prasmės nėra (logikos tarp žodžių), bet gramatiškai teisingas.
- Sakinys nesuprantamas, ir sintaksiškai neteisingas

Šių kriterijų gana daug, kai reikia įvertinti sakinį, todėl dažniausiai pasaulyje vertinimo sistemos vertina pagal kitus kriterijus

- Sakinys yra visiškai teisingas ir gramatiškai taisyklingas
- Sakinys nėra gramatiškai taisyklingas, bet sakinio esmę galima suprasti.
- Sakinio esmės negalima suprasti.

1 Vertimo pavyzdys:

The most critical irony in dealing with Modern Humanism is the inability of its advocates to agree on whether or not this worldview is religious.

Didžiausia kritinė ironija viduje bendravimo su Modern Humanism yra negalėjimas jo gynėjų sutikti bet kuriuo atveju šis pasaulis yra religinis.

Šis sakinytis beveik suprantamas, o ypač aišku, jei žinai kontekstą iš kurio jis paimtas.

2 Vertimo pavyzdys:

Simple sentences:

The brown round table is in a big room. He is a strong man. He is beautiful boy. She is beautiful girl.  
She is a beautiful. He is beautiful.

Paprasti sakiniai:

Rudas apvalus stalas yra dideliame kambaryje. Jis yra stiprus vyras. Jis yra nuostabus berniukas. Ji yra nuostabi mergaitė. Ji yra nuostabi. Jis yra nuostabus.

Štai šie trumpučiai sakiniai yra visiškai teisingi

## 4 Eksperimentinis KV sistemos tyrimas

### 4.1 Eksperimentai su anglų-lietuvių kalbos sistema

Tyrimo dalyvavo KTU humanitarinio fakulteto studentai pasirinkę „Kompiuterinio vertimo“ modulį. Jiems buvo pateikta mūsų sukurta kompiuterinio vertimo programėlė. Kiekvienas studentas turėjo išversti po 3, pusės puslapio apimties tekstus, kur kiekvienas tekstas turėjo būti paimamas iš „Microsoft Word“ dokumentų, „Microsoft Outlook“ ar „Microsoft Outlook Express“ ir parsisiunčiamas iš interneto naudojantis mūsų sukurto sistemos galimybėmis. Jiems taip pat buvo duota paanalizuoti rusų kalbai skirtą programą – PromtX.

Buvo sudarytas testas atsižvelgiant į studentų atsiliepimus apie anglų-lietuvių kalbos, kompiuterinio vertimo (KV) eksperimentinę programą „EiSee“. Studentai bandė ir ne lietuvių kalbai skirtą programą PromtX. Šis testas skirtas numatyti tolesnius KV programos „EiSee“ kūrimo etapus.

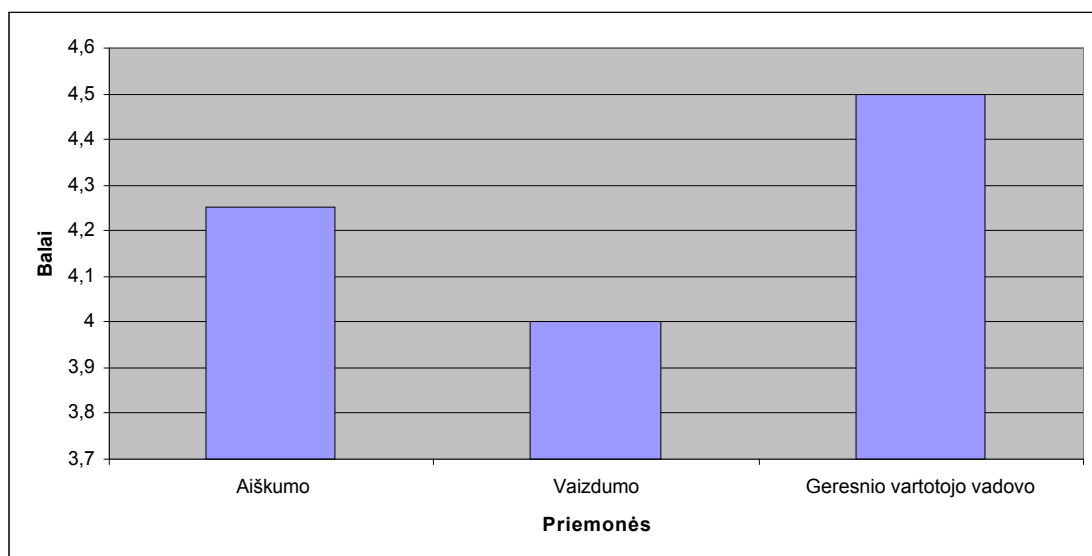
Ko reikia, kad kiekvienas vartotojas lengvai galėtų naudotis programa? Įvertinti jų svarbą. Žiūrėti 23 pav.

**Aiškumo** Pateikta schema kaip veikia programa.

**Vaizdumo.** (Kalbant apie kalbos dalių atributus) Pateikti pavyzdžių. Kokią įtaką tai turės vertimui.

**Vartotojo vadovo programai** Kur ir ką rasti ir kaip naudotis.

Vertinimo sistema 0..5 balų.

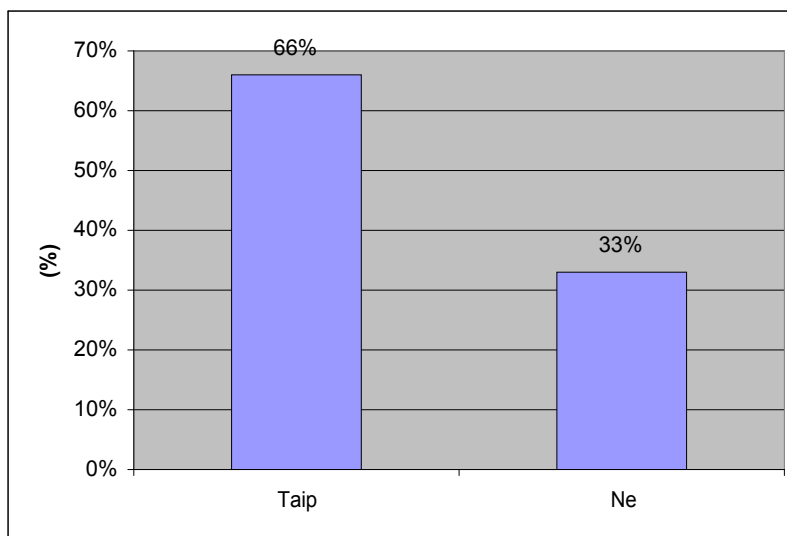


23 pav. Vartotojų poreikių diagrama



Ar leisti paprastam vartotojui įvedinėti žodžius į žodyną bei parinkti atitinkamus atributus tinkamus to žodžio kalbos daliai. Žiūrėti 24 pav.

Vertinti „taip“ arba „ne“



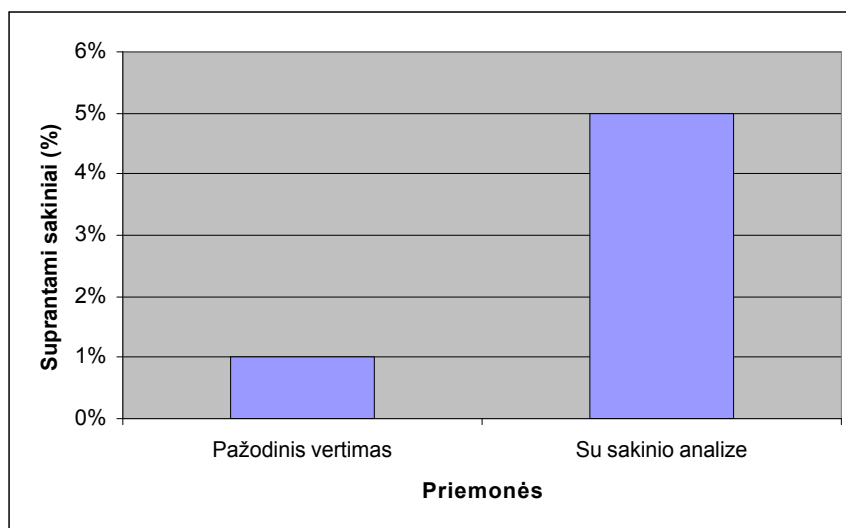
24 pav. Vartotojų nuomonė dėl būtinumo pildyti naujų žodžių atributus

## 4.2 Palyginamieji eksperimentai su anglų-rusų-anglų PROMTX, laalvera ir nauja „EiSee“ vertimo sistema

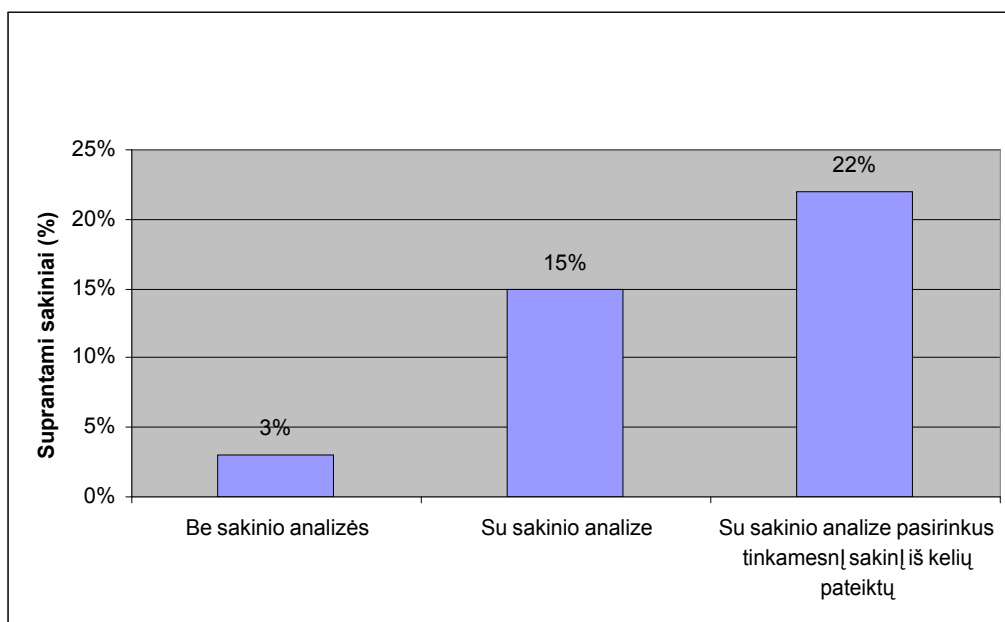
Tyrimė dalyvavo KTU, VDU ir VU įvairių kursų ir specializacijų studentai. Jų buvo paprašyta pasirinkti teksto gabaliuką 10 – 20 sakinių ir pabandyti išversti su „EiSee“ programa panaudojant vertimui režimą, kai programa analizuoja sakinių sintaksę ir semantiką ir pažodinį vertimą. Studentai turėjo įvertinti kiek sakinių buvo:

- Teisingų po pažodinio vertimo (be sakinio analizės).
- Galima suprasti esmę, nors sakiny s nevisiškai taisyklingas naudojantis pažodiniu vertimu (be sakinio analizės).
- Visiškai teisingų (su sakinio analize).
- Galima suprasti esmę, nors sakiny s nevisiškai taisyklingas (su sakinio analize).
- Kiek sakinių buvo galima suprasti po to, kai vartotojas pasirinkdavo tinkamesnį sakinį iš kelių pateiktų (su sakinio analize).

Vartotojai naudojami žodynais kuriuose buvos vos 20 žodžių. Naujus žodžius pildyti reikėjo patiems. Po programos testavimo kiekvienas turėjo 100-200 žodžių savo virtualiuose žodynuose. Rezultatai pateikti 25 ir 26 pav.



25 pav. Visiškai teisingi sakiniai



26 pav. Galima suprasti esmę, nors sakiniai nevisiškai teisingi

Atliktas tyrimas su studentais, atskleidė jog sakinio analizės panaudojimas pagerina vertimo kokybę. Iš 25 ir 26 paveikslų matome, kad vertimo kokybė pagerėja pasinaudojus sakinio analize. Kadangi sakinio analizė pateikiama tikimybiškai, t.y. kiekvienas žodis sakinyje gali būti įvertintas keliomis reikšmėmis, todėl netikslinga griežtai vartotojui nurodyti, kad šis sakinyje verčiamas bus taip, todėl programa sugeneravusi likusius galimus sakinio vertimo variantus, duoda juos vartotojui pasirinkti, tuomet sakinio vertimo kokybė dar pagerėja, nes čia įtraukiama vartotojo nuomonė.

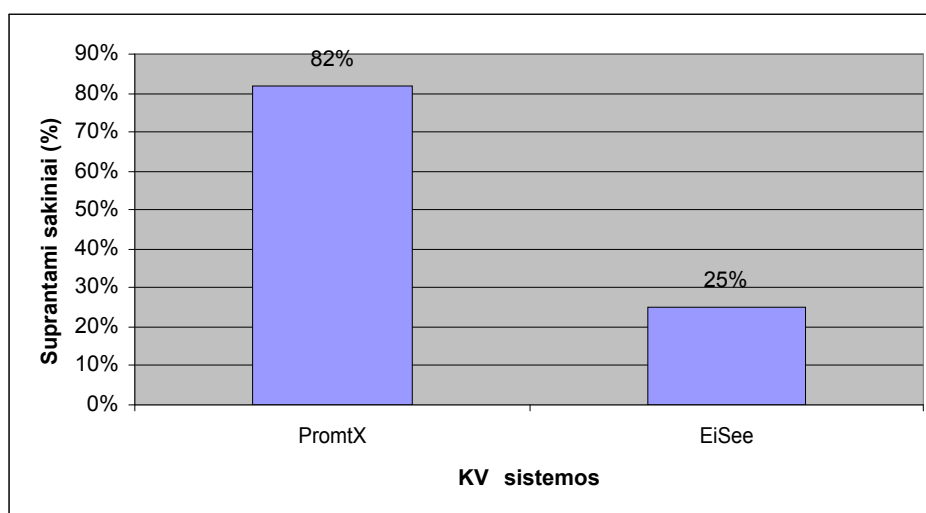
Buvo testuojama rusų-anglų-rusų vertimo sistema PromptX ir lyginama su EiSee programos kokybe 27 pav. EiSee programa naudojami beveik 300 žodžių žodynu.

Programai buvo pateiktas tekstas iš internetinės svetainės [www.yahoo.com](http://www.yahoo.com). Toliau pateikiamas originalus tekstas ir jo vertimas. Abiejų programų versti teksto sakiniai buvo vertinami jei jie buvo suprantami, nepraranda sakinio prasmės.

To serve personal needs, Religious Humanism offers a basis for moral values, an inspiring set of ideals, methods for dealing with life's harsher realities, a rationale for living life joyously, and an overall sense of purpose.

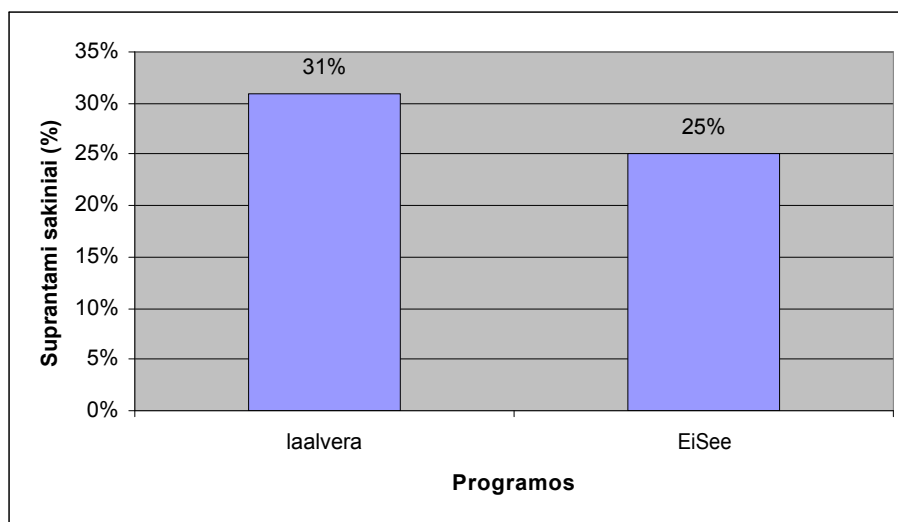
Tarnauti asmeninems naudoms, Religious Humanism siūlo pagrindus moralinei reikšmei, uždegančius kiekius idėjų, metodus bendraujančius su gyvenimų šiurkščiomis tikrovėmis, logiškas gyvenimui džiaugsmingam, ir bendra esmė iš tikslo.

Pilnas tekstas prieduose “KV rezultatas su gramatine sakinių analize”

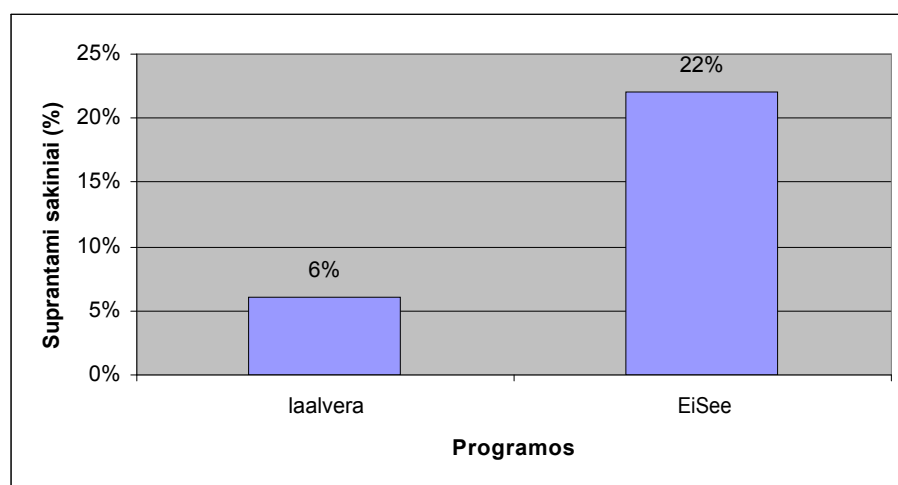


27 pav. Programų kokybės įvertinimas

Kadangi turime ir vieną lietuvių kalbos programą, galime rezultatus palyginti tarpusavyje. Laalvera programa skirta tam tikroms sakinio taisyklėms. Tos taisyklės kaip galima nuspėti apima tik trumpas konstrukcijas (trumpi sakiniai). Programoms buvo pateikiami trumpų sakinių (apie 20) ir po to ilgų sakinių (apie 20). Trumpais sakiniams laikytini 2-4 žodžių, o ilgais laikytini 10-15 ir daugiau žodžių sakiniai.



28 pav. Kokybės įvertinimo diagrama, kai sakiniai trumpi (2-4 žodžiai)



29 pav. Kokybės įvertinimo diagrama, kai sakiniai ilgi (10-15 žodžių)

Laalvera programa daro daug daugiau klaidų, kai pateikiami labai ilgi sakiniai. EiSee programos sakinių suprantamumas truputėli geresnis kai sakiniai trumpesni, bet skiriasi tik 3%

## 5 Išvados

Vartotojų poreikiai gan paprasti: iš šaltinio kalbos išversti į tikslo kalbą, tačiau KV realizavimas yra labai sudėtingas procesas, todėl ją dar labai reikia patobulinti. Sistema yra paruošta darbui su eiliniaisiais vartotojais. Pagrindinis dėmesys ateityje turėtų būti skirtas sakinio analizei ir sakinio analizės susiejimas su žodynų klasifikavimu. Nors programa analizuoja sakinius, pritaiko jiems sintaksės ir semantikos taisykles, tačiau dar šioje srityje dar daug teks padirbėti.

### 5.1 Sistemos realizacija

- Realizuotas sprendimų lentos modelis.
- Realizuota virtualių žodynų koncepcija.
- Įdiegtas frazeologizmų žodynas.
- Specializuotų terminų žodynų valdymas atskirai kiekvienam tekstui.
- Darbas realizuotas plačiai naudojama c++ programavimo kalba.
- Įdiegta šaltinio kalbos sakinio analizė.
- Vertimo į tikslo kalbą analizavimas ir semantinis derinimas.
- Realizuotas vartotojo atmintinės KV sistemos modelis.
- Realizuotas tikslo kalbos sakinio tikslinimo algoritmas, kuris daro tikslo (lietuvių) kalbos sakinius suprantamesniais.
- Suderinta programa su išoriniais žodynais LED ir Alkonas.
- Įdiegtos vartotojo pagalbos priemonės.

### 5.2 Eksperimentų rezultatai

- Eksperimentai parodė, kad vartotojams labai nepatinka ilgai ir nuobodžiai įvedinėti kiekvieno naujo žodžio atributus. Įdiegus sąsają su išoriniais žodynais tokiais kaip „LED“ ir „Alkonas“, naujų žodžių įvedimo laikas sutrumpėjo. Tai pagerino vartotojų nuomonę apie sistemą.
- Įdiegta tikslo kalbos sakinio analizė leidžia gauti geresnių vertimo rezultatų.

- Vartotojams pateikus kelis vieno ir to paties sakinio vertimo variantus, vartotojams daug aiškesnis darosi vertimas. Tokia pagalbos priemonė leidžia pasirinkti gramatiškai teisingesnį sakinį.
- Kiek daugiau nei kas penktas sakinytis gali būti teisingai suprastas (22%). Tokios kokybės sistemų panaudojimas negali būti taikomas publikuojamų tekstų vertimui, o tiksliau tai ji negali būti iš karto publikuojama, būtinas žmogaus priešlaikinis ar rezultato redagavimas.
- Sukurtoji sistema gali būti naudojama anglų kalbos mokymosi tikslais. Jos praktinis panaudojimas labiausiai tinka tose srityse, kur nereikalingas aukštos kokybės vertimas.
- Programa galima naudotis tekstų vertimui naršant internete, kai iš nežinomo teksto norima sužinoti pagrindinę mintį – teksto esmę.
- Plėtojant programą taip, kad vertimas vyktų ir iš lietuvių į anglų, ir iš anglų į lietuvių kalbą, t.y. į abi puses, vartotojo atmintinės modulis neleis kilti dviprasmybėms kai sakinytis bus išverstas iš anglų į lietuvių, o po to tas pats lietuviškas sakinytis bus verčiamas atgal į anglų kalbą.

## 6 Literatūros sąrašas

1. Curme, G.O. (1955) *English Grammar*, N.Y., Barnes and Noble.
2. Garside, R. (1986) *The CLAWS Word - Tagging System*, In: *The Computational Analysis of English*, edited by Garside, R., Leech, G and Sampson, G., Harlow, Longman.
3. Hutchins, W.J. Somers, H. L. *An Introduction to Machine Translation*, [1982]
4. Rosseta, M.T. *Compositional Translation* [1992]
5. Koster, K. *Brief overview of AGFL* [interaktyvus]. 2004, sausis [žiūrėta 2004-01-23].
6. Balkevičius J. Dabartinės lietuvių kalbos sintaksė. V., 1963. [žiūrėta 2004-04-03].
7. Labutis V. Lietuvių kalbos sintaksė. V., 2002. [žiūrėta 2004-03-03].
8. Sirtautas S., Grenda Č. Lietuvių kalbos sintaksė. V., 1988. [žiūrėta 2004-03-03].
9. Specifikacija. Manual for the AGFL system version 2.3  
<http://www.cs.kun.nl/agfl/papers/manual.pdf> [žiūrėta nuo 2004-01-20 iki 2004-04-28]
10. Žodžių tvarka sakinyje [http://ualgiman.tinklapis.lt/zodziu\\_tvarka\\_sakinyje.htm](http://ualgiman.tinklapis.lt/zodziu_tvarka_sakinyje.htm) [žiūrėta 2004-04-25]
11. Sintaksė [http://ualgiman.tinklapis.lt/sintakses\\_savoka.htm](http://ualgiman.tinklapis.lt/sintakses_savoka.htm) [žiūrėta 2004-04-26]
12. Two precursors of machine translation: Artsrouni and Trojanskij John Hutchins  
<http://ourworld.compuserve.com/homepages/WJHutchins/ArtsTroy.pdf> [žiūrėta 2004-04-27]
13. Association for Machine translation and 4th Controlled Language Application Workshop  
<http://www.eamt.org/eamt-claw03>, 2002, Dublin
14. Misevičius, G., Tamulynas, B., Žemaitis, M. Anglų kalbos tekstų kompiuterinio vertimo į lietuvių kalbą technologijos. „Kalbų studijos“/Kaunas: konf. „Informacinės Technologijos 2001“, Kaunas, p. 91-95. 2002. p. 38-45.
15. Sidorov S., Tamulynas B. (2001) Konceptualus hierarchinis kompiuterinio tekstų vertimo modelis// KTU

## 7 Santrauka

Most translation in the world is not of texts which have high literary and cultural status. The great majority of professional translators are employed to satisfy the huge and growing demand for translations of scientific and technical documents, commercial and business transactions, administrative memoranda, legal documentation, instruction manuals, agricultural and medical text books, industrial patents, publicity leaflets, newspaper reports, etc. Some of this work is challenging and difficult. But much of it is tedious and repetitive, while at the same time requiring accuracy and consistency. The demand for such translations is increasing at a rate far beyond the capacity of the translation profession. The assistance of a computer has clear and immediate attractions. The practical usefulness of an MT system is determined ultimately by the quality of its output. But what counts as a 'good' translation, whether produced by human or machine, is an extremely difficult concept to define precisely. Much depends on the particular circumstances in which it is made and the particular recipient for whom it is intended. Fidelity, accuracy, intelligibility, appropriate style and register are all criteria which can be applied, but they remain subjective judgments. What matters in practice, as far as MT is concerned, is how much has to be changed in order to bring output up to a standard acceptable to a human translator or reader. With such a slippery concept as translation, researchers and developers of MT systems can ultimately aspire only to producing translations which are 'useful' in particular situations — which obliges them to define clear research objectives — or, alternatively, they seek suitable applications of the 'translations' which in fact they are able to produce.



## 8 Terminų ir santrumpų žodynas

- **KI** – kalbų inžinerija.
- **KV** – kompiuterinis vertimas.
- **Šaltinio kalba** – kalba, iš kurios verčiami tekstai.
- **Tikslo kalba** – kalba, į kurią norime išversti tekstą.
- **Antraštinis failas** – C++ kalboje naudojami tekstiniai \*.h failai skirti kintamųjų, metodų ir klasių aprašams.
- **API** (Application Programming Interface) – realizuotų operacinės sistemos priemonių (mygtukų, dialogų, lentelių, darbo su procesais ir duomenų srautais priemonių) rinkinys, skirtas tiesioginiam naudojimui rašomose programose.
- **CASE** (Computer Aided Software Engineering) – kompiuterinės programinės priemonės, skirtos palengvinti projektavimą.
- **Direktyva** – instrukcija C/C++ programavimo kalboje nurodanti kompiliatoriui ką reikia atlikti prieš sukompilijuojant programą.
- **DLL** (Dynamic – Link Library) – priemonė leidžianti taikomosioms programoms “užsikrauti” į atmintį šioje bibliotekoje esančius objektus tik esant reikalui. Kuomet objektas, esantis dll, tampa nenaudojamu, jis pašalinamas iš atminties. Tai leidžia taupyti naudojamos atminties kiekį.
- **HTML** (Hyper Text Markup Language) Programavimo kalba skirta kurti internetines svetaines.
- **LKS** – leksinę konceptualiąją struktūrą
- **IDE** (Integrated Development Environment) – Integruota projektavimo aplinka. Paprastai būna specializuota tam tikrai kalbai. Apima programų teksto redaktorių, konkrečios kalbos kompiliatorių, programų derinimo priemones ir pagalbos sistemą. Gali turėti specialias priemones vartotojo aplinkos projektavimui.
- **Noun phrase** - daiktavardžio grupė – daiktavardis ir įeinančios į grupę kalbos dalys: būdvardis, determinantai, artikeliai, įvardis, skaitvardis. *Pastaba:* jei sakinyje iš eilės eina dvi daiktavardžio grupės atskirtos jungtuku arrieveiksmiu, jos gali būti sujungtos į vieną, ištraukiant ir jungiamąjį žodį.
- **Verb phrase** - veiksmažodžio grupė – veiksmažodis ir įeinančios į grupę kalbos dalys: papildantis veiksmažodis anglų kalboje (pvz., “could”, “may”). *Pastaba:* jei sakinyje iš eilės

eina dvi veiksmažodžio grupės, atskirtos jungtuku, jos gali būti sujungtos į vieną, ištraukiant ir jungiamąjį žodį.

- **Preposition phrase** - prielinksnio grupė – prielinksnis.
- **Adverb phrase** - prieveiksmio grupė – prieveiksmis. Jų grupės gali būti sujungtos į vieną, jei jos atskirtos jungtuku.

## 9 Priedai

### 9.1 KV rezultatas su gramatine sakinio analize

The most critical irony in dealing with Modern Humanism is the inability of its advocates to agree on whether or not this worldview is religious.

Those who see it as philosophy are the Secular Humanists while those who see it as religion are Religious Humanists.

This dispute has been going on since the early years of this century when the secular and religious traditions converged and brought Modern Humanism into existence.

Secular and Religious Humanists both share the same worldview and the same basic principles.

This is made evident by the fact that both Secular and Religious Humanists were among the signers of Humanist Manifesto I in 1933 and Humanist Manifesto II in 1973.

From the standpoint of philosophy alone, there is no difference between the two.

It is only in the definition of religion and in the practice of the philosophy that Religious and Secular Humanists effectively disagree.

The definition of religion used by Religious Humanists is a functional one.

To serve personal needs, Religious Humanism offers a basis for moral values, an inspiring set of ideals, methods for dealing with life's harsher realities, a rationale for living life joyously, and an overall sense of purpose.

To serve social needs, Humanist religious communities offer a sense of belonging, an institutional setting for the moral education of children, special holidays shared with like-minded people, a unique ceremonial life.

Religious Humanists maintain that most human beings have personal and social needs that can only be met by religion (taken in the functional sense I just detailed).

They do not feel that one should have to make a choice between meeting these needs in a traditional faith context versus not meeting them at all.

Individuals who cannot feel at home in traditional religion should be able to find a home in non-traditional religion.

I was once asked by a reporter if this functional definition of religion didn't amount to taking away the substance and leaving only the superficial trappings.

My answer was that the true substance of religion is the role it plays in the lives of individuals and the life of the community.

Žemiau pateikiama originalus tekstas ir vertimo rezultatas pasinaudojus AGFL sistema.

The most critical irony in dealing with Modern Humanism is the inability of its advocates to agree on whether or not this worldview is religious.

Didžiausia kritinė ironija viduje bendravimo su Modern Humanism yra negalėjimas jo gynėjų sutikti bet kuriuo atveju šis pasaulis yra religinis.

Those who see it as philosophy are the Secular Humanists while those who see it as religion are Religious Humanists.

Tie kurie gali matyti tai kaip filosofiją yra Secular Humanists kol tie kas mato tai yra religinis yra Religious Humanists.

This dispute has been going on since the early years of this century when the secular and religious traditions converged and brought Modern Humanism into existence.

Šis debatas tęsiasi ant nuo ankščiau metų šio amžium, kai pasaulietinės ir religinės tradicijos konvergavo ir atvėrė Modern Humanism į gyvavimą.

Secular and Religious Humanists both share the same worldview and the same basic principles.

Pasaulietinis ir Religious Humanists kartu dalinasi tą patį pasaulio vaizdą ir tuos pačius pagrindinius principus.

This is made evident by the fact that both Secular and Religious Humanists were among the signers of Humanist Manifesto I in 1933 and Humanist Manifesto II in 1973.

Tai padaryta iš fakto, kad abu Secular ir Religious Humanists among pasirašiusių iš Humanist Manifesto I 1933-aisiais ir Humanist Manifesto II 1973-aisiais.

From the standpoint of philosophy alone, there is no difference between the two.

Iš stovimo filosofijos taško vienas, čia nėra skirtumo tarp du.

It is only in the definition of religion and in the practice of the philosophy that Religious and Secular Humanists effectively disagree.

Tai yra tik aprašyme religijos ir praktikose filosofijos, kad Religious ir Secular Humanists veiksmingiau prieštarauti.

The definition of religion used by Religious Humanists is a functional one.

Religion is that which serves the personal and social needs of a group of people sharing the same philosophical world view.

To serve personal needs, Religious Humanism offers a basis for moral values, an inspiring set of ideals, methods for dealing with life's harsher realities, a rationale for living life joyously, and an overall sense of purpose.

Tarnauti asmeninems naudoms, Religious Humanism siūlo pagrindus moralei reikšmei, uždegančius kiekius idėjų, metodus bendraujančius su gyvenimų šiurkščiomis tikrovėmis, logiškas gyvenimui džiaugsmingam, ir bendra esmė iš tikslo.

To serve social needs, Humanist religious communities offer a sense of belonging, an institutional setting for the moral education of children, special holidays shared with like-minded people, a unique ceremonial life.

Tarnauti socialinėms naudoms, Humanist religiniai komitetai siūlo jausmą priklausomumo, nustatytas rėmas moralei mokslui iš vaikų, ypatingos atostogos dalintis su bendramintiais žmonėmis, unikalų ceremoninį gyvenimą.

Religious Humanists maintain that most human beings have personal and social needs that can only be met by religion (taken in the functional sense I just detailed).

Religinis Humanizmas palaiko kad dauguma žmonių turi asmeniškų ir sociališkų poreikius, kad gali tik būti sutikti religijos (paimtam viduje funkcionalus jausmas I tik aprašiau).

They do not feel that one should have to make a choice between meeting these needs in a traditional faith context versus not meeting them at all.

Jie nejaučia, kad vienas turi turėti šansą tarp susitinkant šiuos poreikius tradicinėje tikėjimą kontekstus prieš nesusitinkant jų iš vis.

Individuals who cannot feel at home in traditional religion should be able to find a home in non-traditional religion.

Individualai kurie nejaučia namuose tradicinėje religijoje turėtų būti galėti rasti namus netradicinėje religijoje.

I was once asked by a reporter if this functional definition of religion didn't amount to taking away the substance and leaving only the superficial trappings.

Aš buvau kartą paklaustas žurnalisto šio funkcinio aprašymo religinio nekiekis paimant šalin medžiaga ir paliekant tik paviršinius atributus.

My answer was that the true substance of religion is the role it plays in the lives of individuals and the life of the community.

Mano atsakymas buvo, kad tiesa medžiaga religijos yra rolė tai žaidžia gyvybėse individualų ir gyvenimas bendruomenės.

## **9.2 Kompiuterinis vertimas naudojantis laalvera programa**

Simple sentences:

The brown round table is in a big room. He is a strong man. He is beautiful boy. She is beautiful girl. She is a beautiful. He is beautiful. They are good people. They could speak. Only little boy can speak. He goes to the school. He learns about dogs. He learns about beautiful dogs. They are beautiful boys. I wrote a very long simple sentence. They went to a beautiful school. He went to a big town. I will go to school tomorrow.

He wrote a letter. They write many letters. I have a big dog. I understand everything. Put it on the table. Table is made of wood.

Pronoun testing:

Give him a box. These men are such awful. This is such a perfect day. Give some flowers for her. She said, that she loves him.

Other sentences:

I like to read books. Do you like to read books? He does not like to read books. Does he like apples? He doesn't like apples. Does he like apples?

I am a boy. And I am a boy. We are boys. You are a boy. She is a girl. He is a boy. They are boys and girls. I'm a good boy. We're good man.

Apple is on the table. You are not a teacher. We are teachers. He is not a teacher. I like big boys. This book is not large. This book is not very large. Where are your parents?

I am not a good teacher.

Sentences with have:

I have a dog. And we have a dog. We have a dog. You have a dog. He has a dog. They have a dog. I've a dog. She hasn't a dog. I have no egg. We don't have lessons on Sunday. We have lessons on Monday. I love the best girl in the world.

Paprasti sakiniai:

Rudas apvalus stalas yra dideliame kambaryje. Jis yra stiprus vyras. Jis yra nuostabus berniukas. Ji yra nuostabi mergaitė. Ji yra nuostabi. Jis yra nuostabus.

Jie esa geri žmonės. Jie galėjo bjauresnis. Tik mažas berniukas gali kalba. Jis eina į mokyklą. Jis mokosi apie šunis. Jis mokosi apie nuostabius šunis.

Jie esa nuostabūs berniukai. Aš rašiau labai ilgą paprastą sakinį. Jie ėjo į nuostabią mokyklą. Jis ėjo į didelį miestą. Aš eisiu į mokyklą rytoj.

Jis rašė laišką. Jie rašo daug laiškų. Aš turiu didelį šunį. Aš suprantu viską. Dėkite tai ant stalo. Stalas yra darė iš medienos.

Įvardžio testavimas:

Duokite jam dėžę. Šie vyrai esa tokio bjaurūs. Šis yra toks puiki diena. Duokite šiek tiek gėlių jai. Ji sakė, kad ji myli jį.

Other sakiniai:

Aš mėgstu skaityti knygas. Ar jūs mėgstate skaityti knygas? Jis nemėgsta skaityti knygų. Ar jis mėgsta obuolius? Jis nemėgsta obuolių. Ar jis mėgsta obuolius?

Aš esu berniukas. Ir Aš esu berniukas. Mes esame berniukai. Jūs esate berniukai. Ji yra mergaitė. Jis yra berniukas. Jie esa berniukai ir mergaitės. Aš esu geras berniukas. Mes esame geri vyrai.

Obuolys yra ant stalo. Jūs esate mokytojai. Mes esame mokytojai. Jis yra mokytojas. Aš mėgstu didelius berniukus. Ši knyga yra dideliė.

Ši knyga yra labai dideliė. Kur esa jūsų tėvai?

Aš esu geras mokytojas.

Sakiniai su turi:

Aš turiu šunį. Ir mes turime šunį. Mes turime šunį. Jūs turite šunį. Jis turi šunį. Jie turi šunį. Aš turiu šunį. Ji turi šunį. Aš turiu ne kiaušinį.

Mes neturime pamokų ant Sekmadienio. Mes turime pamokas ant Pirmadienio. Aš myliu geriausią mergaitę pasaulyje.