

# Programinės įrangos vertimo specifika ir dalinis automatizavimas

Rimgaudas LAUCIUS (MII)

el. paštas: rimga@ktl.mii.lt

**Reziumė.** Straipsnyje nagrinėjama programinės įrangos vertimo specifika bei dalinio jos vertimo automatizavimo priemonių galimybės ir su tuo susijusios problemos. Remiantis statistine verčiamo teksto analize, įvertinamas dažnai pasikartojančių frazių vertimo priemonių taikymo naudingumas.

## 1. Įžanga

Lokalizuojant programinę įrangą išskiriami trys pagrindiniai darbai: 1) adaptavimas lokalei; 2) sąsajos teksto (menui, dialogų langų, pranešimų) vertimas; 3) žinytų vertimas.

Lokalė – tai „naudotojo aplinkos poaibio, priklausančio nuo kalbos ir kultūros normų, apibrėžimas“ (ISO/IEC 15897). Lokalės normoms adaptuota PĮ suteikia galimybes rengti dokumentus, atitinkančius toje lokalėje galiojančias nuostatas bei teisingai vaizduoti išverstą sąsajos tekstą [6].

Adaptavus PĮ lokalės normoms, atliekamas sąsajos teksto ir žinytų vertimas. Nereitai lokalizuojant PĮ apsiribojama jos adaptavimu lokalei ir sąsajos tekstų vertimu, nes žinytų vertimas dėl jų didelės apimties gali būti sunkus ir nemažų investicijų reikalaujantis darbas. Tačiau žinytai taip pat yra svarbi PĮ dalis, todėl bendru atveju, tik atlikus visus tris minėtus darbus, galima laikyti PĮ pilnai lokalizuota

Verčiant žinytus paprastai atliekamas ir sąsajos teksto vertimo testavimas. Verčiant žinytą tenka nuodugnai nagrinėti programą ir paaiškėja didelė dalis programos veikimo ir taikymo aspektų, kurie galėjo likti nepastebėti verčiant sąsajos tekstą, todėl netiksliai išverstas sąsajos teksto frazes galima pataisyti. Kadangi PĮ sąsajos teksto vertimo derinimo sąnaudos yra daug didesnės nei pats vertimas, todėl žinytų vertimas šiuo aspektu taip pat yra gana reikšmingas [5].

Sąsajos ir žinytų teksto vertimas sudaro didžiąją lokalizavimo darbų dalį ir reikalauja nemažų investicijų. Todėl, buvo nuspręsta atlikti tyrimą, siekiant išsiaiškinti ar jo vertimui nebūtų galima panaudoti kokių nors vertimą automatizuojančių priemonių, kurios jį palengvintų.

## 2. Automatizuoto vertimo priemonės

Šiuolaikines automatizuoto vertimo priemones galima suskirstyti į dvi grupes:

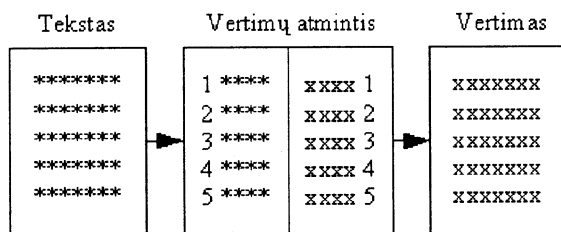
- naudojančios vertimų atmintį (*Translation Memory*);
- naudojančios gramatinę analizę.

Vertimų atmintį EAGLES (Expert Advisory Group on Language Engineering Standards) apibrėžia kaip „daugiakalbio teksto archyvas, talpinantis segmentuotą, rikiuotą, išanalizuotą ir klasifikuotą daugiakalbį tekstą ir sudarantis galimybę laikyti ir išrinkti susietus daugiakalbio teksto segmentus pagal įvairius paieškos kriterijus“. Kitais žodžiais kalbant, tai daugiakalbio teksto duomenų bazė. Šie duomenys yra gana universalūs įvairių vertimo sistemų atžvilgiu, todėl tam, kad jie nebūtų pririšti prie konkrečios programinės įrangos, sukurtas vertimų atminties laikymo ir keitimosi standartas TMX (Translation Memory eXchange). Šis OSCAR (Open Standards for Container/Content Allowing Re-use) atvirasis XML standartas suteikia galimybę keistis vertimų atmintimi tarp skirtingų vertimo programų ir vertėjų su minimaliais svarbios informacijos praradimais.

Vertimų atmintį naudojančių priemonių veikimo principas gana paprastas. Šios priemonės ieško vertimų atmintyje teksto vienetų (sakinijų arba frazių), sutampančių su verčiamo teksto vienetais ir radus juos, automatiškai pakeičia jau anksčiau išverstu tekstu. Šį procesą iliustruoja paveiksle (1 pav.) pateikta schema. Plačiai žinomos šios grupės vertimo priemonės yra: Trados (<http://www.trados.com>), Déjà vu X (<http://www.atril.com>), SDLX (<http://www.sdl.com>), Star TermStar (terminology) and Transit XV (<http://www.star-group.net>), Word Fast (<http://www.champollion.net>), Terminology Wizard (<http://www.synthema.it>), Metatexis (<http://www.metatexis.com>).

Priemonės, naudojančios gramatinę analizę, veikia kitaip. Jos naudoja ne ankstesnių vertimų atmintį, o sudėtingą terminų duomenų bazę, į kurią įtraukti visi gramatiniai kalbos elementai. Šios priemonės versdamos ne tik naudoja išsamius žodynus, tačiau tuo pačiu taiko gramatines taisykles arba duomenų bazėje esančius taisyklių rinkinius tam, kad išversti sakiniai būtų gramatiškai taisyklingi. Šios grupės priemonės gali gana taisyklingai gramatiškai versti tekstą, tačiau jos neišsaugo ankstesnių vertimų. Todėl ateityje verčiant labai panašų tekstą (pvz., tos pačios programos kitos versijos žinyną) jis yra verčiamas nuo nulio.

Skirtingų kalbų gramatika gali būti labai skirtinga, todėl gramatinę analizę naudojančios priemonės nėra universalios skirtingų kalbų atžvilgiu. Kiekvienai skirtingai kalbai yra naudojamas tik jai sukurtas sudėtingas gramatinis analizatorius. Šios grupės priemonių, pritaikytų lietuvių kalbai, kiek autoriui žinoma, dar nėra sukurta. Plačiai žinomos šios grupės priemonės, skirtos užsienio kalboms yra: Prompt Expert (<http://www.prompt.ru>), Personal Translator (<http://www.lingenio.de>), Systran 5.0 (<http://www.systransoft.com>).



1 pav. Vertimo, panaudojant vertimų atmintį, procesas.

Siekdami išsiaiškinti automatinio vertimo priemonių taikymo ypatumus, pabandėme jas taikyti verčiant „OpenOffice.org“ programinę įrangą. Nors priemonės taikytos verčiant konkrečią programinę įrangą, tačiau gauti rezultatai yra bendro pobūdžio.

Visos automatizuoto vertimo priemonės yra gana brangios, todėl bandymui pasirinkome vertimų atmintį naudojančią priemonę Déjà vu X. Ji leidžia bandomojo mėnesio laikotarpiu naudotis visomis jos galimybėmis. Be to, ji geriausiai atitiko mūsų poreikius [1].

Bandymas parodė, jog priemonės, naudojančios vertimų atmintį, iš tiesų gali būti labai sėkmingai taikomos verčiant programinę įrangą. Daugelis frazių nors ir verčiant skirtingą programinę įrangą kartojasi ir todėl, sukauptus nemažą vertimų atmintį, naujų programų arba tų pačių programų naujų versijų vertimą naudojant šias priemones galima atlikti žymiai sparčiau. Tačiau, kol tokia atmintis nėra sukaupta, jų taikymas nėra efektyvus.

Verčiant didelės apimties tekstus, daugelis frazių pasikartoja po kelis ir daugiau kartų tame pačiame tekste. Todėl, vertimo apimtį galima sumažinti išverčiant vieną iš pasikartojančių frazių ir vėliau ją pakeičiant likusias. Mūsų taikyta priemonė Déjà vu X taip pat turi tokią galimybę. Ji išrenka ir parodo dažnai pasikartojančias frazes, tačiau ši jos galimybė mūsų pilnai nepateikino, nes išverstos frazės pakeitimas netgi gana sparčiame kompiuteryje (su 2,7 GHz procesoriumi) užtrukdavo santykinai ilgai. Todėl, šių frazių atrankai buvo nuspręsta paieškoti kitos efektyvesnės ir paprastesnės priemonės.

### 3. Dažnai pasikartojančių teksto frazių paieška

„OpenOffice.org“ olandų lokalizuotojų grupė yra pateikusi dažnai pasikartojančių frazių išskyrimui skirtą priemonę „FrequentStrings“ [2]. Tačiau paaiškėjo, jog ji taip pat nėra tobula: veikia lėtai, naudoja daug operatyviosios atminties, nepateikia jokios informacijos apie verčiamą tekstą ir turi klaidų eliminuojant iš tolesnės paieškos pofrazius. Todėl, teko sukurti naują dažnų frazių išskyrimui skirtą programą. Jai buvo keliami tokie reikalavimai:

- Turi veikti sparčiai ir naudoti santykinai nedaug operatyviosios atminties. „OpenOffice.org“ tekstą sudaro per 400 000 žodžių, o iš jų gaunama per 160 000 frazių, todėl siekiant sukurti sparčiai veikiančią ir santykinai nedaug atminties naudojančią programą, reikia taikyti efektyvius duomenų analizės algoritmus ir programavimo technologijas.
- Turi būti išvengta pofrazių pasikartojimo. Dažnai pasikartojančios frazės nebeturi būti skaičiuojamos ieškant frazių sudarytų iš mažiau žodžių. Turi būti išvengta taip vadinamo „laiptų“ efekto, pavyzdžiui: „Frazė pasikartoja 100 kartų“, „pasikartoja 100 kartų“, „Frazė pasikartoja 100“.

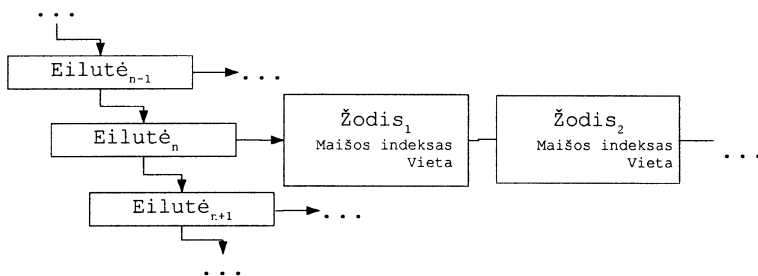
Šias problemas pavyko sėkmingai išspręsti taikant maišos algoritmą ir dinaminio programavimo technologiją. Algoritmo esmę sudaro 3 dalys:

1. Teksto eilutės (po pirminio verčiamo teksto apdorojimo gaunama tekstinė byla sudaryta iš atskirų teksto vienetų – eilučių, tai gali būti ištisi sakiniai, tačiau gali būti ir sakinių frazės) išskaidomos žodžiais ir sukuriamas eilučių dinaminis

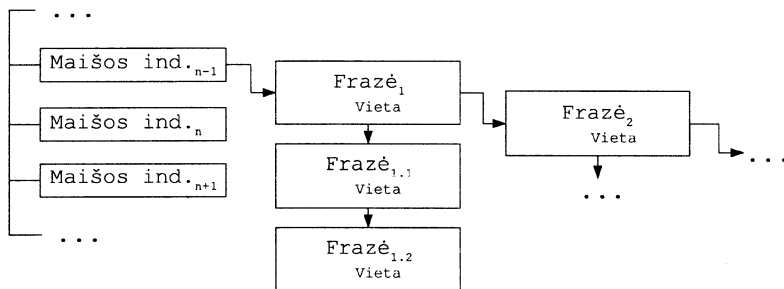
sąrašas, sudarytas iš jas sudarančių žodžių ir jų maišos indeksų masyvų (2 pav.). Tai atlikus, vėliau nebereikia kiekvieną kartą skaičiuoti frazių maišos indeksų iš naujo, o pakanka sudėti frazes sudarančių žodžių maišos indeksus.

2. Iš eilutėse esančių žodžių sudaromos visos galimos frazės iš nustatyto skaičiaus žodžių. Taupant atmintį, frazė žymima nuoroda į eilutę bei pirmo ir paskutinio ją sudarančių žodžių indeksais toje eilutėje. Informacija apie frazę įrašoma į dinaminį sąrašą, kuris įrašomas į masyvo elementą su frazės maišos indeksu (3 pav.). Taip surašius frazes, vėliau jas lengva surasti pagal jų maišos indeksą, todėl vienodų frazių paieška užtrunka labai trumpai.
3. Tikrina frazių masyvą ir išveda į bylą dažnai pasikartojančias frazes bei pašalina jas iš frazių ir žodžių dinaminį sąrašų tam, kad būtų išvengta jų pofrazių pasikartojimo.

Šis algoritmas kartojamas mažinant nustatytą žodžių frazėse skaičių, nuo pasirinktinai ilgiausios frazės iki pasirinktinai trumpiausios frazės. Lyginant šią priemonę su „FrequentStrings“, akivaizdūs šie jos pranašumai: veikia  $\sim 45$  kartus sparčiau, naudoja  $\sim 70$  kartų mažiau operatyviosios atminties, gali pateikti detalią statistinę informaciją apie analizuojama tekstą.



2 pav. Eilučių dinaminis sąrašas sudarytas iš žodžių dinaminį masyvų.



3 pav. Dinaminiai frazių sąrašai.

#### 4. Teksto analizės rezultatai

Siekiant nustatyti kokioms sąlygoms esant patenkinptoms priemonę taikyti yra naudingiausia, išnagrinėjome „OpenOffice.org“ verčiamo teksto sandarą. „OpenOffice.org“ programos sąsajos tekstas savo sandara nuo žinytų teksto beveik nesiskiria. Žinytų tekstas pateikiamas XML formatu ir XML gairės daugumą sakinių suskaido į trumpas frazes, kurios yra būdingos ir sąsajos tekstui. Todėl jų vertimo specifika šiuo atveju artima sąsajos teksto vertimui ir todėl pateiksime tik žinytų teksto analizės rezultatus.

Analizuojant tekstą nustatyta, kad jį sudaro per 75 000 teksto eilučių, kurių vidutinis ilgis 5,4 žodžio (iš viso yra apie 404 350 žodžių, jie sudarytų per 600 mašinraščio puslapių). Ilgiausią eilutę sudaro net 145 žodžiai, tačiau didžiąją teksto dalį (virš 90%) sudaro eilutės trumpesnės nei iš 10 žodžių.

Taip pat buvo nagrinėta atskirų žodžių dažnumo ir jų padengiamo teksto procentais priklausomybė. Nustatyta, kad tekste yra ~2,7% teksto, kuris sudarytas iš nė karto nepasikartojančių žodžių, viso tekste yra 22 460 žodžiai, iš jų 11 083 nepasikartojantys (čia skaičiuojami visi žodžiai, kurie skiriasi bent viena raide, neatsižvelgiant į jų reikšmę). Žodžių, kurie pasikartoja 10 ir daugiau kartų yra 3 524 ir jie padengia ~90% viso teksto; žodžių, kurie pasikartoja 20 ir daugiau kartų – 2 137 ir jie padengia ~85%. Net jei imtume žodžius, kurie pasikartoja 200 ir daugiau kartų, jų yra 273, jie padengtų ~60% teksto.

Dešimt dažniausiai pasikartojančių žodžių yra: *the* 34529 kartus, *to* (9253), *a* (9160), *of* (7919), *in* (7647), *and* (6360), *you* (5518), *is* (4872), *for* (3447), *or* (3410). Kaip ir buvo galima tikėtis, visi šie žodžiai pagalbiniai ir dauguma jų neturi atitikmenų lietuviškame tekste.

Pažvelgus į pateiktus rezultatus gali susidaryti įspūdis, kad pakanka išversti kelis tūkstančius žodžių (maždaug kaip šio straipsnio apimtis) ir likusius automatizuotai pakeikti, vertimas bus beveik atliktas. Tačiau iš tokio pakeitimo geriausiu atveju būtų naudos tik todėl, kad dalies žodžių nereikėtų surinkti nuo pradžios, o juos tik pataisyti, nes kardinaliai skiriasi anglų ir lietuvių kalbų žodžių darybos ir sakinių sintaksės taisyklės. Be to, dauguma angliškų žodžių, tame tarpe ir kompiuterinių terminų, yra daugiareikšmiai. Pavyzdžiui, *From...to* = *Kas...kam* (el. laiške), *Nuo ... iki* (kai kalbama apie intervalą); *check* = *pažymėti* (varnele), *tikrinti* (pvz., rašybą), *tab* = *tabuliacijos ženklas* (dokumente), *kortelė* (dialogo lange) ir t. t. [5].

Nevienareikšmiškumą padaugėja ir dėl to, kad daugelis tų pačių angliškų žodžių (vienodai rašomų) gali atlikti ir daiktavardžio ir veiksmožodžio funkcijas, pavyzdžiui, *file* = *byla, idėti, užpildyti*; *view* = *rodinys, rodyti*; *list* = *sąrašas, išvardyti*; *bookmark* = *adresas, įrašyti adresą* ir t. t. [5].

Dėl šių priežasčių atskirų žodžių vertimas labai priklauso nuo sakinio ar frazės konteksto. Todėl verčiant po vieną žodį gero efekto nepasiektumėme. Gali būti netgi priešingai, tokiu būdu išverstą tekstą būtų žymiai sunkiau derinti ir todėl vertimui reiktų dar didesnių sąnaudų, nei verčiant paprastai. Realią naudą galėtume gauti tik versdami dažnai pasikartojančias frazes, kuriose aiškus kontekstas ir todėl aiškios jų sudarančių žodžių reikšmės bei daryba.

Tačiau kaip nustatyti, kuriose frazėse aiškus kontekstas, o kuriose ne? Suprantama, kuo frazė ilgesnė, tuo didesnė tikimybė, kad jos kontekstas bus aiškus. Tačiau tiksliai nustatyti ilgio ribą nuo kurios frazės arba pakankama frazių dauguma turi kontekstą,

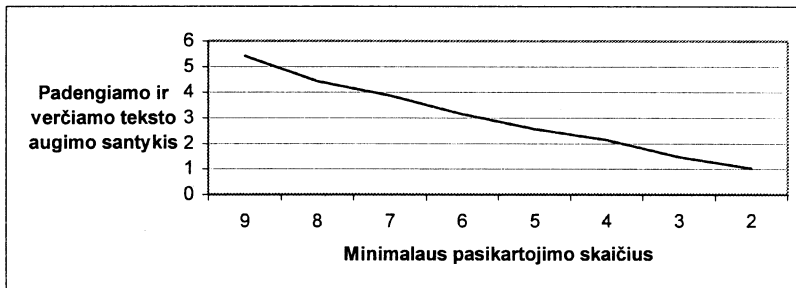
sudėtinga. Iš karto aišku, kad neverta nagrinėti frazių, sudarytų iš dviejų žodžių. Tai pažymi ir kiti tyrinėtojai [7], įvertinę automatinių priemonių taikymo efektyvumą naudodami BLEU metodą [8].

Kita aktuali problema, kurią čia nagrinėsime, – nustatyti minimalų frazių pasikartojimų skaičių, kada iš frazių vertimo automatizavimo būtų galima gauti daugiausia naudos. Akivaizdu, kad mažinant minimalų frazių pasikartojimų skaičių yra padengiama daugiau verčiamo teksto, tačiau dėl to nukenčia vertimo kokybė. Tyrimas parodė, kad maksimalus padengiamo teksto kiekis  $\sim 54\%$  pasiekiamas, kai imamos frazės, kurios pasikartoja po du ir daugiau kartų. Padengiamas tekstas sudaro  $\sim 33\%$ , kai verčiamos frazės pasikartoja 5 ir daugiau kartų ir  $\sim 25\%$ , kai 10 ir daugiau kartų.

Mažinant minimalų frazių pasikartojimų skaičių (PS) padengiama daugiau teksto, tačiau didėja ir iš šių frazių gaunamas verčiamas tekstas. Akivaizdu, kad santykis tarp padengiamo teksto (PT) ir verčiamo teksto (VT) ( $PT/VT$ ) mažėja, o kartu mažėja ir galimo priemonės panaudojimo efektyvumas. Šis mažėjimas yra gana staigus bei artimas tiesinei priklausomybei ( $PT/VT = 2 \times PS - 1,4$ ). Minimumas 2,6 pasiekiamas esant pasikartojimų skaičiui 2 (teoriškai tai reiškia, kad panaudojant priemonę versti reiktų 2,6 kartus mažiau teksto, nei jo padengia frazės) pasiekiamas imant frazes, kurios pasikartoja du ir daugiau kartų.

Tyrimas parodė, kad priklausomybę nuo minimalaus pasikartojimų skaičiaus išreiškiančios funkcijos neturi aiškių lokalių ekstremumų ar linkio taškų, kuriuos galima būtų tyrinėti ir taip nustatyti optimalų minimalaus frazių pasikartojimų skaičių. Iš staigaus padengiamo ir verčiamo teksto santykio mažėjimo aišku, kad padengiamo teksto apimtis mažinant frazių pasikartojimų skaičių auga žymiai lėčiau. Todėl, reikia įvertinti jų augimo santykį ( $PT_{PS+1} - PT_{PS} / VT_{PS+1} - VT_{PS}$ ). Šis santykis pateiktas tolesnėje diagramoje (4 pav.), iš jos matyti, kad esant pasikartojimų skaičiui lygiam dvejetui, santykis artimas vienietui. Tai reiškia, kad nors sumažinę pasikartojimų skaičių nuo trijų iki dviejų, gauname maksimalią padengiamo teksto apimtį, tačiau iš to beveik negauname naudos, nes verčiamo ir padengiamo teksto apimtys padidėja beveik po lygiai.

Iš diagramos matyti, kad mažinant minimalų frazių pasikartojimų skaičių, mažėja priemonės panaudojimo efektyvumas. Be to, žvelgiant į diagramą, taip pat reikia turėti



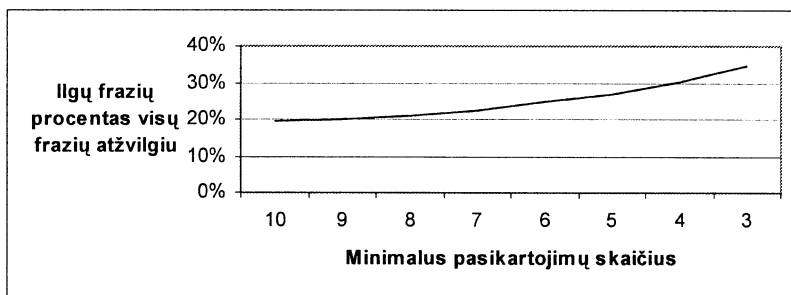
4 pav. Padengiamo teksto ir verčiamo teksto augimo santykio priklausomybė nuo frazių pasikartojimų skaičiaus.

omenyje, kad dalis frazių (ypač trumpų) neturės aiškaus konteksto ir dėl to efektyvumas sumažės dar labiau. Todėl svarbus faktorius, įtakojantis vertimo efektyvumą, – ilgų frazių santykis visų frazių atžvilgiu. Atliekant tyrimą, ilgomis frazėmis laikysime frazes sudarytas iš 5 ir daugiau žodžių. Jų santykio visų frazių atžvilgiu priklausomybė nuo frazių pasikartojimų skaičiaus pateikta diagramoje (5 pav.). Iš jos matyti, kad mažinant frazių pasikartojimų skaičių šis santykis didėja.

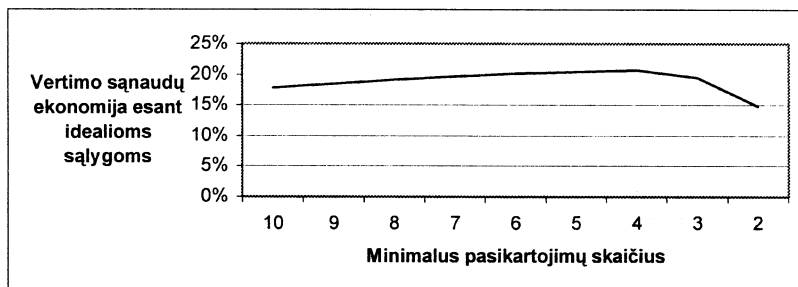
Mažinant minimalų frazių pasikartojimų skaičių didėja padengiamo teksto apimtis, tačiau mažėjantis padengiamo ir verčiamo teksto santykis bei mažėjantis jų augimo santykis rodo automatinio vertimo efektyvumo mažėjimą. Iš dalies ši efektyvumo mažėjimą atsveria didėjantis ilgų frazių santykis visų frazių atžvilgiu, tačiau šis didėjimas nėra didelis ir todėl nėra itin reikšmingas.

Įvertinant padengiamo teksto ir verčiamo teksto santykių viso teksto atžvilgiu esant idealioms sąlygoms, pasiekama vertimo sąnaudų ekonomija procentine išraiška, pateikta diagramoje (6 pav.).

Nors ši diagrama rodo, kad didžiausia ekonomija pasiekama esant pasikartojimų skaičiui lygiam 4, tačiau įvertinant visus nagrinėtus aspektus, rekomenduotinas optimalus frazių pasikartojimų skaičius būtų lygus 5 arba 6. Imant bet kuri iš šių skaičių, ekonomija sumažėtų labai mažai, tačiau priemonės taikymas būtų efektyvesnis dėl didesnio padengiamo ir verčiamo tekstų apimties santykio.



5 pav. Ilgų frazių santykio visų frazių atžvilgiu priklausomybė nuo frazių pasikartojimų skaičiaus.



6 pav. Ilgų frazių santykio visų frazių atžvilgiu priklausomybė nuo frazių pasikartojimų skaičiaus.

Diagramoje pateikta ekonomijos išraiška esant idealioms sąlygoms. Kaip iš jos matyti, esant tokioms sąlygoms maksimaliai vertimo sąnaudas būtų galima sumažinti net penktadaliu. Tačiau dėl jau ankščiau išdėstytų priežasčių, realybėje sąlygos anaip-tol nėra idealios. Be šių priežasčių galima paminėti ir tai, kad panaudojus priemonę, pasunkės išverstų žiny-nų redagavimas. Todėl, įvertinus visas aplinkybes, ekonomija smarkiai sumažėtų. Pagal ankstesnę vertimų patirtį galima apytiksliai numatyti, kad geriausiu atveju taikant priemonę galima tikėtis per pus mažesnės ekonomijos, nei ji būtų pasiekiami esant idealioms sąlygoms.

## 5. Išvados

Vertimo priemonės gali būti labai sėkmingai taikomos verčiant programinę įrangą, kai yra sukaupta pakankama panašaus teksto vertimų atmintis.

Vertimą iš dalies galima automatizuoti verčiant tik po vieną iš dažnai pasikarto-jančių frazių ir vėliau ja pakeičiant likusias. Tokiu būdu galima pasiekti reikšmingą vertimui reikalingų sąnaudų ekonomiją. Pagrindinės priežastys, neleidžiančios pasiekti didelės ekonomijos, yra žodžių daugiareikšmiškumas, žodžių darybos ir sakinių sin-taksės taisyklių skirtumai.

## Literatūra

1. Atril. What's Déjà Vu.  
<http://www.atril.com/whatsdv.html> (žiūrėtas 2004-09).
2. S. Brouwer, Pre-translation of the OpenOffice.org help files.  
<http://nl.openoffice.org> (žiūrėtas 2003-12).
3. EAGLES (Evaluation of Natural Language Processing Systems). Benchmarking translation memories.  
*Doc. EAG-EWG-PR.2*  
<http://issco-www.unige.ch/ewg95> (žiūrėtas 2004-09).
4. G. Dennett, *Translation Memory: Concepts, Products, Impact and Prospects*, South Bank University, London (1995).
5. G. Grigas, Interneto programų paketo lietuvinimo patirtis, *Informacinės technologijos: Konferencijos pranešimų medžiaga*, KTU, Kaunas (2003).
6. R. Laucius, V. Dagienė, Raštinės biuro programinės įrangos „OpenOffice.org“ adaptavimas lokalės normoms. *Konferencijos pranešimų medžiaga. Kompiuterininkų dienos – 2003*, Vilnius (2003).
7. P. Koehn, F.J. Och, D. Marcu, *Statistical Phrase-Based Translation*.  
[www.isi.edu/~marcu/papers/phrases-hlt2003.pdf](http://www.isi.edu/~marcu/papers/phrases-hlt2003.pdf) (žiūrėtas 2004-10).
8. K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a method for automatic evaluation of machine translation. *Technical Report RC22176 (W0109-022)*, IBM Research Report (2001).

## SUMMARY

### *R. Laucius. Software translation specifics and partial automation*

There is analysis of software translation specifics and possibilities of partial automatic translation, presented in this paper. Referencing on statistical text analysis the efficiency of automated translation tools application is estimated.

*Keywords:* automatic software translation, translation memory.