

SAL 30/2017

Tekstynais paremti
vertimų kalbos
tyrimai ir šaltiniai

Received 03/2017

Accepted 05/2017

Tekstynais paremti vertimų kalbos tyrimai ir šaltiniai

Corpus-Based Translation Studies: Research and Resources

TRANSLATION / VERTIMAS

Jurgita Vaičenonienė

Dr., lektorė, Vytauto Didžiojo universitetas, Lietuva.

Jolanta Kovalevskaitė

Dr., mokslo darbuotoja, Vytauto Didžiojo universitetas, Lietuva.

Teresė Ringailienė

Dr., lektorė, Vytauto Didžiojo universitetas, Lietuva.


<http://dx.doi.org/10.5755/j01.sal.0.30.17693>

Anotacija

Tekstynais paremti vertimų kalbos tyrimai yra viena ryškiausių vertimo studijų kryptų. Nors Lietuvoje kai kuriose srityse dominuoja vertimai, o ne originalūs lietuvių kalba rašyti tekstai, vertimų kalba kol kas dar nėra sulaukusi pakankamo Lietuvos tyrėjų dėmesio. Tyrimų stoka gali lemti tai, kad iki šiol neturėta reprezentatyvių lyginamųjų vienakalbių tekstynų, nors tekstynų lingvistikos metodas jau seniai taikomas įvairiems lietuvių kalbos tyrimams. Atsižvelgiant į vertimų mastą rašytinėje lietuvių kalboje šiame straipsnyje siekiama pagrįsti vertimų kalbos tyrimų svarbą Lietuvoje, apžvelgti šia tema atliktus darbus ir pristatyti lyginamąjį vienakalbį tekstyną, sudarytą iš grožinės ir mokslo populiarinimo literatūros tekstų. Straipsnyje remiamasi deskriptyviųjų vertimo studijų bei tekstynų lingvistikos tyrėjų darbais, taip pat paaiškinami pagrindiniai terminai, kai kurių terminų vertimai pateikiami pirmą kartą. Kuriamo lyginamojo originalios ir vertimų lietuvių kalbos tekstyno (ORVELIT) tikslas – reprezentuoti lietuvių vertimų kalbą leksinių–gramatinių ir sintaksės ypatybių tyrimams. Viso tekstyno apimtis apie 4 mln. žodžių; originaliosios ir vertimų kalbos grožinės literatūros ir mokslo populiarinimo kiekvieno iš patekstynių – apie 1 mln. žodžių. Atrenkant tekstus stengtasi subalansuoti tekstyną pagal kūrinio turinio, autorių ir vertėjų įvairovės, lyties, leidyklų ir chronologinių rėmų kriterijus. Tikimasi, kad parengtas šaltinis paskatins gausesnius vertimų kalbos tyrimus, turės mokslinę ir praktinę vertę. ORVELIT lyginamuoju tekstynu bus galima naudotis kaip įrankiu, padedančiu suvokti tarpžanrinius tekstų reikalavimus ir ypatumus, atlikti tam tikrų leksinių vienetų paiešką, įvertinti vertimų ir originalių tekstų skirtumus.

REIKŠMINIAI ŽODŽIAI: lietuvių kalba, vertimų kalba, lyginamieji vienakalbiai tekstynai, tekstynais paremti vertimo tyrimai, vertimo universalijos, ORVELIT.



Vertimo studijose teigiama, kad vertimai ne tik daro įtaką kalbai, bet ir yra savarankiška jos atmaina. Vertimai ir originalūs tos pačios kalbos tekstai gali būti suprantami kaip skirtingi registrai su jiems būdingomis ypatybėmis (Baker, 1993, 1995). Šis požiūris turėtų būti ypač aktualus mažiau vartojamoms kalboms, kurios yra nuolat veikiamos dominuojančių kalbų. Lietuvai po Nepriklausomybės atgavimo tokią įtaką daro anglų kalba. Iš šios kalbos verčiama daugiausia, o kai kuriose srityse iš anglų kalbos verčiama daugiau nei publikuojama lietuviškų tekstų. Pavyzdžiui, 2014-ųjų metų Lietuvos spaudos statistikos duomenimis, iš anglų kalbos išverstos knygos sudarė net 63 % visų vertimų. Be to, vertimų tiražas vidutiniškai dvigubai didesnis (1,9 tūkst. egz.) nei originalių knygų (1 tūkst. egz.), o tai rodo didesnę vertimų paklausą tarp skaitytojų; didžioji dalis vertimų (net 60 %) yra grožinės literatūros kūriniai (Markevičienė ir Tamulygienė, 2015).

Lietuviškuose naujienų portaluose (pvz., užsienio, mokslo, pramogų, sporto ar kitų naujienų rubrikose) ir informaciniuose, mokslo populiarinimo bei pramoginiuose periodikos leidiniuose (pvz., *A-ZET*, *National Geographic Lietuva*, *Cosmopolitan*, *Iliustruotasis mokslas* ir kt.) nemažai informacijos yra verčiama iš įvairių užsienio šaltinių. Taigi galima daryti prielaidą, kad Lietuvoje grožinės literatūros ir mokslo populiarinimo leidiniuose vertimų pasiūla yra didesnė už originalią lietuvišką kūrybą. Būtent pramoginę grožinę ir įvairių sričių mokslo populiarinimo literatūrą lietuviškų leidyklų redaktoriai bei bibliotekų darbuotojai mini kaip daugiausia skaitomą (*Lietuvių skaitymo ypatumai*, 2015). Pateikti pavyzdžiai ne tik parodo vertimų ir originaliosios lietuvių kalbos santykį, bet ir pagrindžia šių dviejų kalbos atmainų lyginamųjų tyrimų poreikį ir svarbą.

Vertimų kalbos tyrimuose daugiausia dėmesio iki šiol skiriama didžiosioms kalboms, tokioms kaip anglų (Baker, 1993; Laviosa, 1997), vokiečių (Kenny, 2001) ar kinų (Xiao et al., 2010), tačiau vis dar neaišku, ar didžiosiose kalbose nustatytos vertimų kalbos tendencijos būdingos ir tipologiškai skirtingoms mažosioms kalboms, pvz., lietuvių. Tekstynais paremti vertimų kalbos tyrimai yra viena ryškiausių kryptių moderniose vertimo studijose. Kita vertus, ši studijų ir tyrimų kryptis kol kas dar nėra sulaukusi pakankamo Lietuvos tyrėjų dėmesio (Vaičėnienė, 2011; 2014). Tyrimų stoką gali lemti tai, kad, nors tekstynų lingvistikos metodai jau seniai taikomi įvairiems lietuvių kalbos tyrimams, iki šiol neturėta pakankamai reprezentatyvių lyginamųjų vienakalbių tekstynų. Daugiakalbiuose lygiagrečiuosiuose tekstynuose reprezentuojama lietuvių kalba yra vertimų kalba, todėl trūko palyginamojo tekstyno, kuris galėtų reprezentuoti ir vertimų, ir originalią lietuvių kalbą.

Atsižvelgiant į vertimų mastą rašytinėje lietuvių kalboje šiame straipsnyje siekiama trumpai aptarti tekstynais paremtų vertimo studijų raidą ir šaltinius, naudojamus vertimo kalbai tirti, vertimų kalbos tyrimus Lietuvoje. Straipsnyje taip pat pristatomas parengtas vienakalbis lyginamasis tekstynas, sudarytas iš grožinės ir mokslo populiarinimo literatūros tekstų (*Lyginamasis originalios ir vertimų lietuvių kalbos tekstynas, toliau – ORVELIT*). Straipsnyje remiamasi deskriptyviųjų vertimo studijų bei tekstynų lingvistikos tyrėjų darbais, taip pat paaiškinami pagrindiniai terminai, kai kurių terminų vertimai pateikiami pirmą kartą.

Vertimo studijų skiriamos dvi kryptys – teorinė (angl. *Pure Translation Studies*) ir taikomoji (angl. *Applied Translation Studies*) (Malmkjaer, 2012, p.31). Teoriniams tyrimams priskiriamos deskriptyviosios vertimo studijos (angl. *Descriptive Translation Studies*), kur atsiribojama nuo preskriptyviojo vertimų vertinimo ir skatinami sisteminiai, empirija paremti vertimų tyrimai (Touy, 1995; Malmkjaer, 2012). Deskriptyviųjų vertimo studijų šakai priskiriamos ir tekstynais paremtos vertimo studijos (angl. *Corpus Based Translation Studies*). Šioje dalyje bandysime trumpai apžvelgti šios krypties raidą, užsienio ir lietuvių autorių darbus ir esamus šaltinius.

Įvadas

Tekstynais paremti vertimų kalbos tyrimai ir šaltiniai

Vertimų kalbos tyrimai: vertimų universalijos

Kitaip nei tradicinėse vertimo paradigmos, deskriptyviuosiuose vertimo tyrimuose nagrinėjamas ne vertimo procesas, kai lyginamas originalas ir vertimas, bet analizuojamas ir apibūdinamas vertimo proceso rezultatas (Toury, 1995; Malmkjaer, 2012). Toks požiūris gali padėti išvengti išankstinių nuostatų dėl idealų, lūkesčių (ne)atitinkančių ekvivalentiškumo aspektų ir leidžia analizuoti vertimus kaip savarankiškus tekstus (Toury, 1995; Malmkjaer, 2012, p.34). Šios G. Toury'io (1995; 1995) idėjos paskatino lyginamuosius įvairių žanrų vertimų bei pačios vertimų kalbos tyrimus (Malmkjaer, 2012, p.34). Nors vertimų kalbos tyrimų būta ir anksčiau (pvz., Vanderauwera, 1985; Blum-Kulka, 1986), M. Baker (1993, 1995, 1996) juos ne tik apibendrino, bet ir viena pirmųjų pasiūlė tiriant vertimų kalbą taikyti tekstynų lingvistikos metodologiją, padėsiančią nustatyti vertimų kalbos ypatybes.

M. Baker teigimu, vertimai ne tik skiriasi nuo originalių tekstų, bet ir pasižymi universalio- mis ypatybėmis (Baker, 1993). Egzistuoja tam tikros **vertimo kalbos universalijos** – kalbos ypatybės, dažniau pasitaikančios verstuose, o ne originaliuose tekstuose, ir šios ypatybės nepriklauso nuo kalbos, iš kurios verčiama (Baker, 1993, p.243). M. Baker (1996) pasiūlė keturias tikėtinas vertimų kalbos universalijas, kurias galima rasti lyginamuosiuose viena- kalbiuose įvairių kalbų tekstynuose: paprastinimas (angl. *simplification*), arba polinkis var- toti skurdesnę kalbą; eksplicitiškumas (angl. *explicitation*), arba polinkis į daugiažodžiamumą; norminimas (angl. *normalization*), arba polinkis paprastinti tai, kas vertimo kalbai neįprasta, perdėtas vertimo kalbos normų laikymasis; neutralizavimas (angl. *levelling out*), arba polin- kis vengti stilistiškai žymėtos raiškos.

Nepaisant įvairių diskusijų dėl universalijų hierarchijos, tikslumo ir apibrėžčių, per pasta- ruosius dvidešimt metų M. Baker idėjos ypač išpopuliarėjo ir dabar plėtojamos daugelio ty- rėjų darbuose, siekiant atrasti įvairių kalbos ypatybių, kurios leistų pagrįsti vertimų kalbos universalijas (Laviosa, 1997, 1998; Kenny, 2001; Zanettin, 2013). Organizuojama tarptautinių konferencijų šia tema (pavyzdžiui, Tarptautinės vertimo raštu ir kultūrinių studijų asociacijos (IATIS) 2004, 2006, 2009, 2012, 2015 m. konferencijų ciklas; R. Xiao organizuotos konferen- cijos, skirtos tekstynų taikymams lyginamosiose ir vertimo studijose (UCCTS 2010, 2012, 2013, 2014). Metodologiškai vertimų kalbos tyrimuose dažniausiai naudojami vienakalbiai lyginamieji ir dvikalbiai lygiagretieji ir (ar) lyginamieji tekstynai. Tyrėjai pasirenka vieną ar kelias kalbos ypatybes, kurių pasiskirstymo, dažnumo ir vartojimo palyginimas atskleidžia tos pačios kalbos originalių ir verstų tekstų panašumus bei skirtumus. Dominuoja mažos apimties, t. y., vienos ypatybės, žanro ar universalijos tyrimai, o kompleksinių, didelės apim- ties, lyginamųjų darbų nėra daug (pvz., Xiao et al., 2010).

Vis dėlto iš turimų tyrimų galima spręsti, kad vertimo proceso metu vertėjai iš tikrųjų są- moningai ir (ar) nesąmoningai paprastina kalbą, neperteikdami kūrybiškos raiškos, yra linkę laikytis tipinių vertimo kalbai raiškos formų ar įterpti tikslinančios informacijos (Zanettin, 2012, p.13). Toliau aptarsime keletą dažniausia minimų darbų.

Atliekant lyginamuosius įvairių kalbų porų tyrimus ypač dažnai remiamasi S. Laviosos (1997, 1998) anglų vertimų kalbos tyrimais ir metodologija. Lygindama anglų vertimų kalbos teks- tyno (angl. *Translational English Corpus*) ir Britų nacionalinio tekstyno (angl. *British National Corpus*) duomenis, S. Laviosa ištyrė, kad angliškuose vertimuose **paprastinimas** matyti iš mažesnės leksinės įvairovės, tankio ir informacinio svorio:

- _ santykis tarp visų ir skirtingų teksto žodžių yra mažesnis nei originaliuose tekstuose;
- _ dažnai vartojamų žodžių sąrašas yra ilgesnis, negu retai vartojamų žodžių;
- _ dažni žodžiai vertimuose vartojami dažniau nei originaluose, taip pat turi mažiau lemy;
- _ prasminių žodžių mažiau nei tarnybinių;
- _ vidutinis sakinio ilgis yra trumpesnis (nepasitvirtinusi hipotezė).

Kinų kalbos tyrimas (Xiao et al., 2010) taip pat atskleidė, kad versti kinų kalbos tekstai pasižymi mažesne leksine įvairove. Kadangi kinų ir anglų kalbos tipologiškai labai tolimos, šie rezultatai ypač svarūs siekiant pagrįsti paprastinimo universaliją.

Eksplacitiškumo tyrimai (Olohan ir Baker, 2000) rodo, kad vertimams būdingos perteklinės sintaksinės struktūros, kurios vertimo kalboje paprastai būna pasirenkamos, o ne privalomos. Pavyzdžiui, pasirenkamas jungtukas *that* (liet. *kad*) po veiksmazodžių *say* ir *tell* (liet. *sakyti, pasakyti*) dažnesnis vertimų nei originalioje anglų kalboje. Vertimuose į ispanų kalbą įvardžiai subjekto pozicijoje yra dažnesni, o originalioje ispanų kalboje – retesni (Jiménez-Crespo, 2011). Kita vertus, kai kurių tyrėjų nuomone, eksplacitiškumas turėtų būti suprantamas ne kaip vertimo universalija, o kaip originalo kalbos ypatybių perkėlimas ir įtaka vertimo kalbai (plg. Saldanha, 2008).

Norminimo, dar kitaip vadinamo standartizavimo (Toury, 1995), tyrimais siekiama išsiaiškinti, ar įprasta, konvencinė, o ne kūrybinė raiška vertimuose pasirenkama dėl to, kad siekiama nenukrypti nuo vertimo kalbos normų. Remdamasi vokiečių literatūros vertimų į anglų kalbą tekstynu D. Kenny (2001) tikrino hipotezę, kad vertimuose neperteikiama kūrybinė raiška. Tyrimo rezultatai parodė, kad originale pavartoti kūrybiški žodžiai dažnai verčiami įprastais vertimo kalbai žodžiais ir frazėmis. Kad leksiką linkstama versti neišlaikant originalo kūrybiškumo, rodo ir neįprastų kolokacijų vertimo analizė: galima matyti gana aiškų vertėjų polinkį norminti neįprastus žodžių junginius.

M. Baker pasiūlyta **neutralizavimo** universalija nesusilaukė tyrėjų dėmesio. Kita vertus, atsirado naujų vertimo universalijų siūlymų, pavyzdžiui, S. Trikkonen-Condit (2002) ir S. Eskola (2004) suformuluota **unikalių kalbos vienetų stokos hipotezė** (angl. *translation of unique items*). Autorių teigimu, kalbos vienetai, kurie egzistuoja vertimo, o ne originalo kalboje, vertimams nebūdingi. Tyrimai parodė, kad suomių kalbai būdingi leksiniai, sintaksiniai ir pragmatiniai konstruktai vertimuose iš anglų ir rusų kalbų pasitaikė rečiau nei originaliuose suomių kalbos tekstuose. K. Malmkjær (2008, p.57) teigia, kad būtent ši hipotezė galėtų būti laikoma perspektyviausia, nes atspindi kognityvinius sprendimus, o ne vertimo kalbos normų interferenciją, ir gali padėti pagrįsti, jog vertimų kalbos universalijos egzistuoja.

Universalijų tyrimų populiarumas lėmė ir terminų problemas: ta pati universalija gali būti suprantama plačiau ar siauriau, egzistuoja keli tą patį reiškiniį įvardijantys terminai, siūlomos naujos universalijos (pvz., perkėlimas (angl. *transfer*); unikalių kalbos vienetų stoka (angl. *unique items*); netipinių kolokacijų vertimas (angl. *untypical collocations*); interferencija (angl. *interference*). F. Zanettinas (2012, 2013) parodo, kiek skirtingų pavadinimų yra vartojama toms pačioms universalijoms: pavyzdžiui, norminimą (angl. *normalization*) kiti autoriai vadina standartizavimu (angl. *standardization*), konservatizmu (angl. *conservatism*), konvencionalizacija (angl. *conventionalization*); eksplacitiškumas gali būti vadinamas asimetrija (angl. *asymmetry*). Nepaisant terminų variacijos ir diskusijų apie vertimų universalumą ar kaitą, priklausančią nuo tikslinės kalbos ir kultūros normų, tyrėjai sutinka, kad vertimo kalba turi tik jai būdingų, savitų ypatybių, o šis požiūris skatina kuo įvairesnių kalbų porų ir krypčių tyrimus ir sudėtingesnių, anotuotų tekstynų kūrimą.

Per pastaruosius dvidešimt metų Europoje ir JAV sukurta daugiau kaip šimtas įvairių tekstynų, skirtų vertimų tyrimams (Hu, 2016, p.9). Toliau apibendrinsime tik kai kuriuos dažniau minimus (plg. Zanettin, 2012; Kruger et al., 2013; Hu, 2016) vertimų kalbos tyrimų šaltinius:

1 Atviros prieigos, vienakalbis, vienakryptis anglų vertimų kalbos tekstynas (*Translational English Corpus (TEC)*), pradėtas kurti 1996 m. vadovaujant M. Baker. 10 mln. žodžių tekstyną sudaro angliški grožinės literatūros, laikraščių kalbos, biografiniai ir skrydžio žurnalų tekstai, versti iš įvairių kalbų. Naudotas paprastinimo (pvz., Laviosa, 1997; 1998), eksplacitiškumo (Olohan ir Baker, 2000), standartizavimo (Olohan, 2004) universalijų tyrimams.

Vertimų
kalbos
tekstynai

- 2 Lygiagretusis, dvikryptis, morfologiškai anotuotas 2,6 mln. žodžių grožinės ir negrožinės literatūros anglų–norvegų kalbų tekstynas (*The English–Norwegian Parallel Corpus (ENPC)*), parengtas Oslo universitete (1994–1997) vadovaujant S. Johanssonui. 1997–2001 m. pridėtos vokiečių, olandų ir portugalų kalbos (*Daugiakalbis Oslo tekstynas (Oslo Multilingual Corpus (OMC))*). Atlikti eksplicitiškumo (Øverås, 1998) tyrimai. Prieiga galima tik universiteto tyrėjams.
- 3 Vienakalbis, lyginamasis suomių vertimų kalbos tekstynas (*The Corpus of Translational Finnish (CTF)*), sudarytas vadovaujant A. Mauraneen (2000). Tekstyną (9,6 mln. žodžių) sudaro originali suomių ir versta grožinė, vaikų, akademinė ir mokslo populiarinimo literatūra. Tyrinėtos paprastinimo (Mauraneen, 2000), unikalių kalbos vienetų (Trikkonen-Condit, 2002; Eskola, 2004) universalijos. Prieigos nėra.
- 4 Dorothy Kenny (2001) sudarytas dvikalbis, vienakryptis, lygiagretusis, šiuolaikinės grožinės vokiečių literatūros vertimų į anglų kalbą tekstynas (*German English Parallel Corpus of Literary Texts (GEPCOLT)*). Naudotas norminimo universalijos tyrimui. Tekstyno apimtis maždaug 2 mln. žodžių. Atviros prieigos nėra, neanotuotas.
- 5 Lygiagretusis, dvikryptis, 4,6 mln. žodžių, grožinės ir negrožinės literatūros anglų–italų vertimų tekstynas (*English–Italian Translational Corpus (CEXI)*), rengtas Bolonijos universitete, Forli, vadovaujant G. Aston. Tekstynas kurtas remiantis anglų–norvegų kalbų tekstyno (*The English–Norwegian Parallel Corpus (ENPC)*) sandaros principais. Projektas taip ir nebuvo baigtas nutrūkus finansavimui (Zanettin, 2012).
- 6 CroCo projekte (*Linguistic Properties of Translations. A Corpus-Based Investigation for the Language Pair English–German (CROCO)*), 2005–2010 m. vykdytame Saarlando universitete (Hansen-Schirra et al., 2006) parengtas anotuotas, dvikalbis, dvikryptis, lygiagretusis, aštuonių skirtingų registrų anglų–vokiečių kalbų tekstynas. 1 mln. žodžių tekstynas anotuotas gramatiškai, daugiausia tirtas vertimų eksplicitiškumo požiūriu (pvz., Hansen-Schirra et al., 2007).
- 7 Lankasterio universitete sudarytas lyginamasis kinų kalbos tekstynas (*the Lancaster Corpus of Mandarin Chinese (LCMC)*), kinų vertimų kalbos tekstynas (Xiao, 2010). Tekstyno apimtis – 1 mln. žodžių, reprezentuojami įvairūs žanrai, anotuotas, laisvai prieinamas internete. Daug dėmesio skirta paprastinimo ir eksplicitiškumo tyrimams (Xiao et al., 2010).
- 8 Lygiagretusis, dvikryptis, anotuotas, tęstinis, 3 mln. žodžių portugalų–anglų kalbų originalų ir vertimų tekstynas (*COMPARA: Portuguese – English Parallel Translation Corpus*), sukurtas pagal anglų–norvegų kalbų tekstyno modelį (Frankenberg-Garcia et al., 2003). Tyrinėtos paprastinimo ir eksplicitiškumo universalijos (pvz., Frankenberg-Garcia, 2009). Laisvai prieinamas. Be pavyzdžių, kurie jau išvardyti, dar kuriami lygiagretieji daugiakalbiai ar lyginamieji tekstynai kitoms kalbų poroms, pavyzdžiui, prancūzų–ispanų, anglų–rusų, anglų–slovėnų, anglų–lenkų, etc. Nors egzistuoja nemažai resursų vertimų kalbos tyrimams, F. Zanettinas (2012, p.23) pastebi, jog susiduriama su tekstynų palyginamumo problemomis; mažosios kalbos ne visuomet turi pakankamai originalių ar verstų tam tikrų žanrų tekstų, kad būtų įmanoma atitikti didžiųjų kalbų tekstynų apimtis ir standartus; reikėtų daugiau dėmesio skirti natūralios kalbos apdorojimo technologijoms ir tekstynų anotavimui; trūksta pakartotinių, lyginamųjų skirtingų kalbų porų ir žanrų, tarpdisciplininių tyrimų.

**Tekstynais
paremti
vertimų tyrimai
Lietuvoje**

Lietuvoje taip pat gausėja vertimo studijoms skirtų darbų, kuriuose taikomi tekstynų ir kompiuterinės lingvistikos metodai. Galima skirti keletą tyrimų kryptų. Visų pirma tai straipsniai, pristatantys įvairius šaltinius, tekstynus, jų kūrimo ir analizės principus, programinę įrangą (pvz., Marcinkevičienė, 2002; Mikelionienė, 2002; Utkā, 2004; Rimkutė et al., 2006, 2007, 2008; Utkā et al., 2012). Žinoma, šių darbų pritaikomumas apima daugelį skirtingų mokslo sričių, bet taip pat ypač svarbus ir galimiems vertimų kalbos tyrimams.

Pavyzdžiui, VDU Kompiuterinės lingvistikos centre sukurti ir viešai prieinami lietuvių–latvių ir latvių–lietuvių kalbų lygiagretusis tekstynas (LILA) (Utka et al., 2012); dar anksčiau – dvikalbis lygiagretusis anglų–lietuvių, čekų–lietuvių kalbų tekstynas (Rimkutė et al., 2008), sudarytas iš tokių patekstynių:

- | | |
|--|---|
| – Anglų–lietuvių (70 813 lygiagretūs sakiniai; 2 023 005 žodžiai), | – Čekų–lietuvių (41 881 lygiagretūs sakiny; 821 479 žodžiai), |
| – Lietuvių–anglų (1614 lygiagrečių sakinių; 61 154 žodžiai), | – Lietuvių–čekų (693 lygiagretūs sakiniai; 21 064 žodžiai). |

Projektas nėra baigtinis ir ateityje tekstyną planuojama pildyti naujomis kalbų poromis ir tekstais.

Antra, tai tyrimai, kuriuose tekstynai ir tekstynų lingvistikos metodai taikomi analizuojant tarpkalbinius raiškos skirtumus, vertimų atitikmenis leksikos, gramatikos, sintaksės, teksto rišlumo ar pragmatikos lygmenimis (Kalėdaitė, 2006; Ruzaitė, 2010; Vaičenonienė, 2012). Naudojami jau sukurti, atviros prieigos, arba kuriami konkreta tyrimo tikslams įgyvendinti reikalingi tekstynai. Pavyzdžiui, S. Volungevičienės (2013) tyrimui apie kolokacijų vertimo iš vokiečių į lietuvių kalbą problemas sukurtas lygiagretusis vokiečių–lietuvių kalbų tekstų ir jų vertimų tekstynas (1 mln. žodžių), sudarytas iš įvairių kanceliarinių, administracinių, grožinės literatūros ir populiarinio mokslo kūrinių bei jų vertimų. A. Šolienės (2013) disertacijoje „Episteminio modalumo ekvivalentiškumo parametrai anglų ir lietuvių kalbose“ duomenys išrinkti iš specialiai sukurtos dvikalbio dvikrypčio lygiagrečiojo anglų–lietuvių kalbų tekstyno (1 572 498 mln. žodžių), sudaryto iš originalių grožinės literatūros tekstų ir jų vertimų.

Trečia, mažiau tyrėjų dėmesio sulaukusi, tekstynais paremtų vertimų kalbos tyrimų kryptis. Pavyzdžiui, J. Vaičenonienės (2011) disertacijoje nagrinėjama leksikos standartizacija (norminimas) vertimuose. Sukurtame lygiagrečiajame grožinės lietuvių literatūros vertimų į anglų kalbą tekstyne (1 690 164 mln., žodžių, 759 841 lygiagrečių sakinių) išrinkti ir analizuoti 166 stilistiškai žymėti okaziniai ir potencialiai žodžiai ir jų vertimai. Tyrimo rezultatai atskleidė, kad net 64 % autorinių neologizmų vertimuose neišliko. Kita vertus, vertimo strategijų įvairovė ir vertėjų požiūrio į tą patį vertimo vienetą skirtumai rodo, jog standartizacija yra daugiau tam tikra norma, kintanti priklausomai nuo kultūrinių, kalbinių ar kitų faktorių, o ne universali vertimo kalbos ypatybė (plg. Kenny, 2001). Kitame tyrime (Vaičenonienė, 2014) aiškinamasi, kaip skiriasi originalų ir vertimų kalba neanotuotame lyginamajame grožinės literatūros tekste, sudarytame iš *Dabartinės lietuvių kalbos tekstyno* ir *Lygiagrečiojo* tekstyno duomenų. Originalios lietuvių literatūros patekstynio imtis – 6 721 812, o vertimų iš anglų kalbos patekstynio – 2 976 425 žodžiai. Žodžių dažnumo ir įvairovės, žodžių sąrašų, prasminių ir tarnybinių žodžių pasiskirstymas parodė, kad vertimų kalbos leksika yra ne tokia įvairi, jai būdingas paprastinimas. Įvardžių vartosenos analizė remia sintaksinio eksplacitiškumo universalijos idėją – vertimuose kai kurie įvardžiai dažnesni negu originaliuose lietuviškuose kūriniuose.

Nors kuriami lygiagretūs ir lyginamieji tekstynai, šių tekstynų apimtis gana ribota, dalis jų viešai neprieinami, neanotuoti. Tekstynų kūrimo kriterijai skiriasi, o tai neleidžia atlikti lyginamųjų tyrimų. Dominuoja grožinės literatūros vertimų tekstynai ir tyrimai, ypač iš (j) anglų kalbą, trūksta tarpžanrinių studijų. Kita vertus, kaip jau minėta anksčiau, tokios duomenų gausos, nesuderinamumo, prieigos ir saugojimo problemos bendros ir kitų šalių vertimo studijoms, ieškoma būdų joms spręsti (Zanettin, 2012).

Lyginamieji tekstynai gali būti vienkaliai, dvikalbiai ar daugiakaliai, jų lyginamąsias dalis sudaro tais pačiais kriterijais (pvz., teksto dydžio, turinio, laikotarpio ir kt.) atrinkti tekstai, vienodai ar panašiai reprezentuojantys pasirinktą kalbos atmainą vienoje ar keliose kalbose (Olohan, 2004; Kenning, 2010). Pavyzdžiui, gali būti kuriami lyginamieji daugiakaliai lai-

kraščių, politikos, darbo skelbimų tekstynai (Kenning, 2010, p.487) arba vienakalbiai originalų ir vertimų tekstynai (Olohan, 2004, p.35). ORVELIT yra vienakalbių lyginamųjų tekstynų tipui priskirtinas tekstynas.

M. Nelsonas (2010, p.54) atkreipia dėmesį, kad prieš kuriant naują tekstyną, svarbu įvertinti jo poreikį, naujumą ir svarbą. Kadangi lyginamojo vertimų ir originalų tekstyno lietuvių kalbai iki šiol nebuvo, ORVELIT poreikis yra pagrįstas. Kuriamo tekstyno tikslas – reprezentuoti lietuvių vertimų kalbą leksinių–gramatinių ir sintaksės ypatybių tyrimams. Tekstyno sandara leis tirti vertimo kalbos ypatybes dviejuose, grožinės literatūros ir mokslo populiarinimo, registruose. Svarbu pabrėžti, kad šiuo tekstynu siekiama atspindėti vertimų iš anglų kalbos ypatumus, o ne reprezentuoti lietuvių grožinę ar mokslo populiarinimo literatūrą apskritai.

Kuriant ORVELIT tekstyną vadovautasi principais, kurie pateikiami duomenų tvarkybos planuose (angl. *Data Management Plan*) ir yra svarbūs kuriamų kalbos išteklių ilgalaikiškumui, sklaidai ir prieinamumui užtikrinti. Duomenų tvarkybos planuose išskiriami tokie žingsniai:

1 Duomenų rinkimo sprendimai (pvz., duomenų pobūdis, kiekis, prieinamumas, saugojimo formatas, ilgalaikiškumas, taikomi standartai, metodologijos, metaduomenų standartai ir kt.).

2 Dokumentacijos ir metaduomenų sprendimai (duomenų rinkimo, klasifikacijos, pobūdžio ir t. t. aprašymas reikalingas ateities tyrimams; egzistuojančių metaduomenų

aprašymo standartų taikymas/ adaptavimas – tyrimo tikslams).

3 Etikos ir teisiniai sprendimai (duomenų rinkimo, saugojimo, platinimo, anonimiškumo užtikrinimo, autorinių teisių klausimai).

4 Pareigos ir resursai (išteklių kūrimo ir saugojimo biudžetas, darbų paskirstymas ir kt.) (Trippel ir Zinn, 2015, p.71).

ORVELIT tekstynui taikytas Europos mokslinių tyrimų infrastruktūros humanitariniams ir socialiniams mokslams, skirtos kalbos ištekliams (duomenims ir įrankiams) CLARIN ERIC siūlomas duomenų tvarkybos plano vediklis (*CLARIN-D*).

Kuriant vertimų kalbos tekstynus vadovaujamosi panašiais kūrimo principais kaip ir sudarant bet kuriuos kitus tekstynus (Zanettin, 2012, p.47). Tam, kad būtų atspindėtos vertimų kalbos ypatybės, tekstynas turėtų būti pakankamai didelis, subalansuotas, patekstynius sudarytų pakankamai tekstų, kurie reprezentuotų kalbų, autorių, leidėjų, autorių ir vertėjų įvairovę (Zanettin, 2012, pp.47–48). Toliau aptarsime grožinės literatūros ir mokslo populiarinimo patekstynių sandarą.

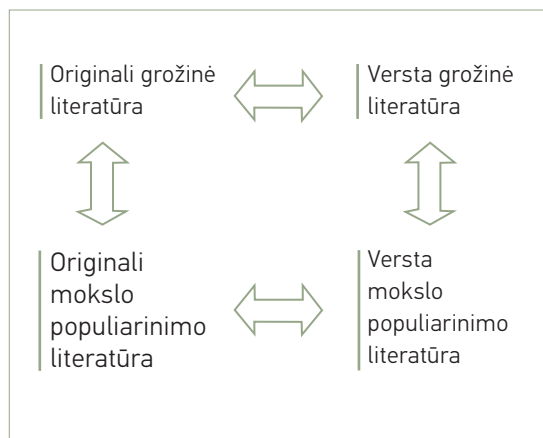
ORVELIT susideda iš keturių patekstynių, kurių ryšiai gali būti vaizduojami 1 paveiksle.

Teoriškai grožinės literatūros vertimų patekstynį galima būtų lyginti su esamo *Dabartinės lietuvių kalbos tekstyno* (DLKT) grožinės literatūros patekstyniu, kuris atspindi skirtingų laikotarpių, kartų, lyčių, žanrų ir kt. literatūrą.

Kita vertus, kadangi vertimų literatūros patekstynio turinys daug siauresnis, kyla klausimas, ar palyginimas būtų validus. Be to, DLKT grožinės literatūros dalyje gali būti tiek originalių, tiek verstų kūrinių, rašytų gimtakalbių ir negimtakalbių, išėivijos atstovų, įtraukti atsiminimai ar kiti žanrai. DLKT yra ir negrožinės literatūros dalis, kurioje be mokslo populiarinimo tekstų yra ir mokslo, mokomosios, pažintinės literatūros tekstai, taigi ši dalis taip pat negali būti laikoma palyginama mokslo populiarinimo pa-

1 paveikslas

ORVELIT patekstynių palyginamumas



tekstynei. Dėl to nuspręsta iš DLKT esančių duomenų sukurti atskirus originalios grožinės ir mokslo populiarinimo literatūros patekstynius, kurie savo sandara būtų kuo panašesni į vertimų kalbos patekstynius ir reprezentuotų ne grožinę ir mokslo populiarinimo literatūrą apskritai, bet originalią lietuvių kalbą.

Viso *ORVELIT* tekstyno apimtis – 4 034 347 mln. žodžių; originaliosios ir vertimų kalbos grožinės literatūros ir mokslo populiarinimo patekstynių – daugiau nei 1 mln. žodžių. Tokio dydžio imtis laikoma pakankama tyrinėjant didelio dažnumo gramatines, leksines ypatybes ir specializuotus žanrus (Evison, 2010, p.123). Tekstyną sudaro kūriniai (visateksčiai), išsaugoti .txt formatu, ANSI (*American National Standards Institute*) koduote. Žodžių skaičiui nustatyti naudota *WordSmith Tools 7.0* (Scott, 2016) versija. Atrenkant tekstus stengtasi subalansuoti tekstyną pagal kūrinių turinio, autorių ir vertėjų įvairovės, lyties, leidyklų ir chronologinių rėmų kriterijus (žr. 1 lentelę). Kai kurie sprendimai, pavyzdžiui, dėl chronologinių tekstyno

Kriterijus	Grožinės literatūros vertimų patekstynis	Originalios grožinės literatūros patekstynis	Mokslo populiarinimo vertimų patekstynis	Originalios mokslo populiarinimo literatūros patekstynis
<i>Tipas</i>	Specializuotas, lyginamasis, vienakalbis, vienakryptis			
<i>Reprezentatyvumas</i>	Anglų kalbos grožinės literatūros vertimų kalba	Lietuvių grožinės literatūros kalba	Anglų kalbos mokslo populiarinimo tekstų vertimų kalba	Lietuvių mokslo populiarinimo tekstų kalba
<i>Populiacija</i>	Grožinė literatūra, versta iš anglų kalbos	Originali grožinė lietuvių literatūra	Mokslo populiarinimo literatūra, versta iš anglų kalbos	Originali lietuvių kalba rašyta mokslo populiarinimo literatūra
<i>Imties rėmai</i>	<i>Dabartinės lietuvių kalbos tekstyno</i> grožinės literatūros dalis; <i>Lygiagrečiojo</i> tekstyno vertimų dalis	<i>Dabartinės lietuvių kalbos tekstyno</i> grožinės lietuvių literatūros dalies kūriniai, rašyti lietuvių autorių	<i>Dabartinės lietuvių kalbos tekstyno</i> negrožinės literatūros dalis; <i>Lygiagrečiojo</i> tekstyno vertimų dalis	<i>Dabartinės lietuvių kalbos tekstyno</i> negrožinės literatūros dalies kūriniai, rašyti lietuvių autorių
<i>Imties vienetas</i>	Knyga tekstinio failo formatu			
<i>Žodžių skaičius</i>	1 003 747 mln. žodžių	1 003 519 mln. žodžių	1 010 830 mln. žodžių	1 016 251 mln. žodžių
<i>Tekstų skaičius</i>	18 visateksčių kūrinių	18 visateksčių kūrinių	19 visateksčių kūrinių	25 visateksčiai kūriniai
<i>Pobūdis</i>	Rimtoji ir populiarioji literatūra	Rimtoji ir populiarioji literatūra	Humanitarinių, socialinių, ekonomikos mokslų, medicinos populiarinimas	Humanitarinių, socialinių, gamtos ir ekonomikos, informacinių technologijų, medicinos mokslų populiarinimas
<i>Autoriai</i>	18 skirtingų autorių	18 skirtingų autorių	19 skirtingų autorių	24 skirtingi autoriai
<i>Lytis</i>	8 moterys, 10 vyrų	9 moterys, 9 vyrai	7 moterys, 16 vyrų	15 moterų, 19 vyrų
<i>Vertėjai</i>	15 skirtingų vertėjų	-	19 skirtingų vertėjų	-
<i>Leidyklos</i>	12 skirtingų leidyklų	5 skirtingos leidyklos	6 skirtingos leidyklos	10 skirtingų leidyklų
<i>Chronologiniai vertimų rėmai</i>	1981–2011	-	2005–2012	-
<i>Chronologiniai kūrinių rėmai</i>	1952–2006	1994–2012	1963–2011	2005–2012

1 lentelė

Vertimų ir originalų patekstynių sandara

rėmų, buvo nulemti ribotos *Lygiagrečiojo* ir *Dabartinės lietuvių kalbos* tekstynų imties, t. y., norint išlaikyti pakankamą autorių, vertėjų ir temų įvairovę, nuspręsta įtraukti ne tik šiuolaikinę literatūrą, bet ir kūrinius, rašytus nuo XX a. vidurio.

Pastebėjome, kad ypač problemiškas kriterijus yra lyčių balansas: nors grožinės literatūros patekstyne vyrų ir moterų autorių skaičius – apylygis, mokslo populiarinimo patekstyne taikyti šį kriterijų buvo sudėtingiau, nes kūrinių pasirinkimo galimybės buvo ribotesnės: imties rėmuose prieinami kūriniai turėjo atitikti mokslo populiarinimo žanro apibrėžimą, o tai susiaurino ir temų įvairovę.

Pats mokslo populiarinimas gali būti suprantamas keleriopai. Anot Burns ir kt. (2003), tai gali būti ir visuomenės susidomėjimas mokslu, o kartais šis terminas vartojamas tiesiog kalbant, kaip visuomenė supranta mokslą, mokslinę kultūrą ir raštingumą. Sudarant mokslo populiarinimo patekstinį, laikytasi pačios bendriausios sampratos: įvairių žanrų tekstai, kuriuose specializuotas mokslas transformuojamas taip, kad jį suprastų nespecialistai; kitaip tariant, mokslu siekiama sudominti nespecialistus (Calsamiglia ir van Dijk, 2004).

Taigi dėl mokslo populiarinimo patekstylio laikytasi nuostatos, kad lyties kriterijus ne toks svarbus kaip temų įvairovė. Ir mokslo populiarinimo vertimų, ir originaliosios mokslo populiarinimo literatūros patekstinį sudaro humanitarinių, socialinių, ekonomikos, medicinos mokslų populiarinimo tekstai. Kai kurios sritys, pavyzdžiui, informacinių technologijų ar gamtos mokslų, būdingos tik originaliai mokslo populiarinimo daliai. Kitaip nei grožinės literatūros patekstyne, mokslo populiarinimo dalyje kai kurie tekstai rašyti dviejų ar daugiau autorių, o tai irgi atskleidžia skirtingų žanrų ypatumus. Nors tekstyne ir nepavyko išlaikyti vienodo balanso tarp visų išskirtų atrankos kriterijų, manome, kad tai nėra didelis trūkumas. Viena vertus, esama sudėtis atskleidžia realią rašomų tekstų situaciją, kita vertus, tekstynai nebūtinai yra baigtiniai ar išaldyti. Kiekvienas tyrėjas gali modeliuoti ir koreguoti tekstyno turinį ir apimtį pagal savo tyrimo tikslus.

Pirmoji tekstyno versija, planuojama, bus saugoma Lietuvos nacionalinio konsorciumo CLARIN-LT¹ centro saugykloje. Prieiga prie tekstyno bus galima su riboto naudojimo pobūdžio (RES) licencija, kuri apibrėžia naudotojo tapatybės ir prieigos sąlygas, bendrąsias naudojimo ir platinimo sąlygas (*Teisinė CLARIN-LT veiklos studija*, 2015, pp.29–30). Be visateksčių dokumentų, tyrėjams gali būti pateikiami ir tokie duomenys kaip dažninių žodžių sąrašai, metaduomenys ir kita tyrimams aktuali informacija.

Sudarant tekstyną buvo vadovautasi teisinės CLARIN-LT veiklos studijoje (2015) pateiktu išaiškinimu dėl kuriamų išteklių autorinių teisių, kurioje teigiama, kad CLARIN-LT saugomi duomenys, o tai reiškia ir tekstynai, nepažeidžia autorinių teisių, numatytų pagal Lietuvos Respublikos autorinių teisių ir gretutinių teisių įstatymą:

CLARIN-LT turinys – autorinio turinio nedidelės dalys nurodant šaltinį, kūriniai (net visateksčiai) moksliniais arba informaciniais tikslais – gali būti naudojami be originalaus autoriaus (teisių turėtojo) žinios, be leidimo ir be atskiro atlyginimo (*Teisinė CLARIN-LT veiklos studija*, 2015, p.23). Sudarant ORVELIT tekstyną visi tekstai buvo pernaudoti iš *DLKT* ir *Lygiagrečiojo* tekstynų, dėl to nereikėjo iš naujo spręsti autorinių teisių klausimų.

¹ CLARIN-LT yra Bendrosios kalbos išteklių ir technologijų infrastruktūros CLARIN ERIC B tipo centras. Viena CLARIN-LT siūlomų paslaugų yra duomenų saugojimas, kuriuo siekiama užtikrinti šaltinių, tekstynų, programinės įrangos ir kitokių duomenų ilgaamžiškumą, prieigą, vienodus socialinių ir humanitarinių mokslų duomenų kūrimo standartus ir kokybę.

Nors atlikta nemažai tyrimų ir rasta įvairių vertimo kalbos ypatybes liudijančių įrodymų, daugiausia tiriamos didžiosios kalbos, todėl ypač skatinami ir reikalingi mažiau paplitusių, įvairresnių kalbų ir jų porų tyrimai. Taip pat reikėtų pastebėti, kad dominuoja literatūros vertimų tyrimai, tuo tarpu kitiems žanrams ar jų lyginimui skiriama mažiau dėmesio. Taigi tarpžanrinis vertimo ypatybių tyrimas leistų objektyviau įvertinti jų paplitimą ir universalumą.

Manoma, kad sukurto naujo tekstyno ORVELIT, kurį ateityje numatoma morfologiškai ir sintaksiškai anotuoti bei suteikti atvirą prieigą, tyrimų rezultatai turės ne tik mokslinę, bet ir praktinę vertę ruošiant profesionalius įvairių sričių vertėjus, filologus. ORVELIT lyginamuoju tekstynu bus galima naudotis kaip įrankiu, padedančiu suvokti tarpžanrinius tekstų reikalavimus ir ypatumus, atlikti tam tikrų leksinių vienetų paiešką, įvertinti vertimų ir originalių tekstų skirtumus. Tekstynas bus naudingas ir vertimo studijų ar filologijos dėstytojams kaip pagalbinė priemonė ruošiant mokymo užduotis. Gautos išvados apie vertimų kalbos ypatybes palengvins profesionalių vertėjų, kalbos redaktorių darbą, padės objektyviai vertinti vertimų kokybę. Taip pat tikimasi, kad bus paskatinti tolimesni tyrimai, įtraukiantys vertimus iš skirtingų kalbų grupių, analizuojantys specializuotų originalių ir verstų tekstų, pavyzdžiui, teisinės ar techninės kalbos, ypatumus, kuriams ne tik lyginamieji, bet ir lygiagretūs ar multimodalūs tekstynai, skirti ne tik rašytinei, bet ir sakytinei vertimo kalbai.

Apibendrinimas

Literatūra

1. Advokatų profesinė bendrija FORT, 2015. Teisinė CLARIN-LT veiklos studija. Vilnius. [abs/10.1177/09636625030122004](https://doi.org/10.1177/09636625030122004) [Žiūrėta 2016 m. gruodžio mėn].
2. Baker, M., 1993. *Corpus Linguistics and Translation Studies: Implications and Applications*. Francis, G., and Tognini-Bonelli, E., eds., *Text and Technology: Honour of John Sinclair* Amsterdam/ Philadelphia: John Benjamins, pp.233–250.
3. Baker, M., 1995. *Corpora in Translation Studies. An Overview and Some Suggestions for Future Research*. *Target*, 7 (2), pp.223–243. <https://doi.org/10.1075/target.7.2.03bak>
4. Baker, M., 1996. *Corpus-based Translation Studies – the Challenges that Lie Ahead*. In: Somers, H., ed., *Terminology, LSP and Translation*. Philadelphia/ Amsterdam: John Benjamins, pp.175–86. <https://doi.org/10.1075/btl.18.17bak>
5. Blum-Kulka, S., 1986. *Shifts of Cohesion and Coherence in Translation*. In: House, J., Blum-Kulka, S., *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition*. Tübingen: Günter Narr Verlag, pp.17–35.
6. Burns, T. W., O'Connor, D. J., Stocklmayer, S. M., 2003. *Science Communication: A Contemporary Definition*. *Public Understanding of Science* 12 (183), pp.183–202. Prieiga per internetą: <http://journals.sagepub.com/doi/abs/10.1177/09636625030122004> [Žiūrėta 2016 m. gruodžio mėn].
7. Calsamiglia, H., van Dijk, T., 2004. *Popularization Discourse and Knowledge about the Genome*. *Discourse and Society* 15 (4), pp.369–389. <https://doi.org/10.1177/0957926504043705>
8. Eskola, S., 2004. *Untypical Frequencies in Translated Language: A Corpus-based Study on Literary Corpus of Translated and Non Translated Finnish*. In: Mauraneen, A., Kujamäki, eds., *Translation Universals. Do they Exist?* Amsterdam/ Philadelphia: John Benjamins, pp.83–99. <https://doi.org/10.1075/btl.48.08esk>
9. Evison, J., 2010. *What are the Basics of Analysing a Corpus?* In: O'Keeffe, A., McCarthy, M., *The Routledge Handbook of Corpus Linguistics*. London/ New York: Routledge, pp.122–135. <https://doi.org/10.4324/9780203856949.ch10>
10. Frankenberg-Garcia, A., Santos, D., 2003. *Introducing COMPARA, the Portuguese–English Parallel Translation Corpus*. In: Zanettin, F., Bernardini, S., Stewart, D., eds., *Corpora in Translation Education*. Manchester: St. Jerome Publishing, pp.71–87. Prieiga per internetą: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.7050&rep=rep1&type=pdf> [Žiūrėta 2016 m. liepos mėn.]

11. Frankenberg-Garcia, A., 2009. Are Translations Longer than Source Texts? A Corpus-based Study of Explicitation. In: Beeby, A., Rodriguez, P., Sanchez-Gijon, P. (eds.), *Corpus Use and Translating*. Amsterdam/ Philadelphia: John Benjamins, pp.47–58. <https://doi.org/10.1075/btl.82.05fra>
12. Hansen-Schirra, S., Neumann, S., Vela, M., 2006. Multi-dimensional Annotation and Alignment in an English–German Translation Corpus. *Proceedings of the Workshop on Multi-dimensional Markup in Natural Language Processing, EACL*, pp.35–42. <https://doi.org/10.3115/1621034.1621040>
13. Hansen-Schirra, S., Neumann, S., Steiner, E., 2007. Cohesive Explicitness and Explicitation in an English–German Translation Corpus. *Information Structuring Resources in Contrast, Languages and Contrast*, 7(2), pp.241–265.
14. Hu, K., 2016. *Introducing Corpus-based Translation Studies*. Heidelberg/ New York: Springer, p. 9. <https://doi.org/10.1007/978-3-662-48218-6>
15. International Association for Translation and Cultural Studies. Prieiga per internetą: <http://www.iatis.org/index.php/iatis-belo-horizonte-conference> [Žiūrėta 2016 m. liepos mėn.]
16. Jiménez-Crespo, M. A., 2011. The Future of General Tendencies in Translation: Explicitation in Web Localization. *Target*, 23(1), pp.3–25. <https://doi.org/10.1075/target.23.1.01jim>
17. Kalėdaitė, V., 2006. *Translating Existence (A Corpus-based Analysis)*. *Darbai ir Dienos*, 45, pp.119–125.
18. Kenning, M. M., 2010. What are Parallel and Comparable Corpora and How Can We Use Them? In: O’Keeffe, A., McCarthy, M., eds. *The Routledge Handbook of Corpus Linguistics*. London/ New York: Routledge, pp.487–500. <https://doi.org/10.4324/9780203856949.ch35>
19. Kenny, D., 2001. *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester: St Jerome.
20. Kruger, A., Wallmach, K., Munday, J., 2013. *Corpus-Based Translation Studies. Research and Applications*. London/ New York: Bloomsbury.
21. Laviosa, S., 1997. How Comparable Can Comparable Corpora Be? *Target*, 9(2), pp.289–319. <https://doi.org/10.1075/target.9.2.05lav>
22. Laviosa, S., 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* 43 (4), pp.557–570. <https://doi.org/10.7202/003425ar>
23. Lietuvių skaitymo ypatumai, 2015. Veidas. [prieiga internetu] <http://www.veidas.lt/lietuviu-skaitymo-ypatumai>. [Žiūrėta 2016 m. liepos mėn.]
24. Lietuvos Respublikos autorinių teisių ir gretutinių teisių įstatymas. Prieiga per internetą: <https://www.etar.lt/portal/lt/legalAct/TAR.551F0CDE5B64/UMOLkJpCuL> [Žiūrėta 2016 m. liepos mėn.]
25. Malmkjaer, K., 2008. Norms and Nature in Translation Studies. In: Anderman, G., Rogers, M., eds., *Incorporating Corpora – Corpora and the Translator*. Clevedon: Multilingual Matters, pp.49–59.
26. Malmkjaer, K., 2012. Where Are We? (From Holme’s Map until Now). *The Routledge Handbook of Translation Studies*. London/ New York: Routledge, pp.31–44.
27. Marcinkevičienė, R., 2002. Palyginamieji tekstynai – šaltinis tarptautinių žodžių vartosenai tirti. *Kalbotyra*, 51 (3), pp.81–93.
28. Markevičienė, R., Tamulygienė, L., 2015. Lietuvos spaudos statistika. Vilnius: Lietuvos nacionalinė Martyno Mažvydo biblioteka.
29. Mauraneen, A., 2000. Strange Strings in Translated Language. A Study on Corpora. In: Olohan, M.(ed.) *Intercultural Faultlines*. Manchester: St Jerome Publishing, pp.119–141.
30. Mikelionienė, J., 2002. Palyginamojo tekstyno kūrimo principai, problemos ir panaudojimo galimybės. *Kalbų studijos*, 3, Kaunas: Technologija, pp.55–59.
31. Nelson, M., 2010. *Building a Written Corpus. What are the Basics?* *The Routledge Handbook of Corpus Linguistics*. London/ New York: Routledge, pp.53–65.
32. Olohan, M., Baker, M., 2000. Reporting That in Translated English: Evidence for Subconscious Processes of Explicitation? *Across Languages and Cultures*, 1(2), pp.141–158. <https://doi.org/10.1556/Acr.1.2000.2.1>
33. Olohan, M., 2004. *Introducing Corpora in Translation Studies*, London/ New York: Routledge.

34. Øverås, L., 1998. In Search of the Third Code: An Investigation of Norms in Literary Translation. *Meta*, 43(4), pp.557–570. <https://doi.org/10.7202/003775ar>
35. Rimkutė, E., Kovalevskaitė, J., Daudaravičius, V., 2006. Daugiakalbių tekstynų naudojimas ir taikymas. *Darbai ir Dienos*, 45, pp.41–62.
36. Rimkutė, E., Daudaravičius, V., 2007. Morfologinis Dabartinės lietuvių kalbos tekstyno» anotavimas. *Kalbų studijos*, 11, pp.30–35.
37. Rimkutė, E., Daudaravičius, V., Utkā, A., Kovalevskaitė, J., 2008. Bilingual Parallel Corpora for English, Czech and Lithuanian. *Proceedings of The Third Baltic Conference on Human Language Technologies 2007*, pp.319–326.
38. Ruzaitė, J., 2010. Translation Equivalents of Vague Language Items: a Study of General Extenders in a Parallel Corpus. *Kalbų studijos*, 16, pp.33–38.
39. Saldanha, G., 2008. Explicitation Revisited: Bringing the Reader into the Picture. In: House, J., *Beyond Intervention: Universals in Translation Process*, trans-kom 1(1), pp.20–35. Prieiga per internetą: http://www.trans-kom.eu/bd01nr01/trans-kom_01_01_03_Saldanha_Explicitation.20080707.pdf [Žiūrėta 2016 m. liepos mėn.].
40. Scott, M., 2016. WrodsSmith Tools. Prieiga per internetą: <http://www.lexically.net/wordsmith/> [Žiūrėta 2016 07 31].
41. Šolienė, A., 2013. Episteminio modalumo ekvivalentiškumo parametrai anglų ir lietuvių kalbose. (Daktaro disertacija). Vilnius: VU.
42. Toury, G., 1995. *Descriptive Translation Studies and Beyond*. Amsterdam/ Philadelphia: John Benjamins. <https://doi.org/10.1075/btl.4>
43. Trikkonen-Condit, S., 2002. Translationese – a Myth or an Empirical Fact. A Study into the Linguistic Identifiability of Translated Language. *Target*, 14(2), pp.145–171.
44. Trippel, T., and Zinn, C., 2015. DMPTY – A Wizard for Generating Data Management Plans. *CLARIN 2015 Selected Papers*. Linköping Electronic Conference Proceedings, no. 123. Prieiga per internetą: <http://www.ep.liu.se/ecp/123/006/ecp15123006.pdf> [Žiūrėta 2016 m. liepos mėn.].
45. Utkā, A. 2004. Phases of Translation Corpus: Compilation and Analysis. *International Journal of CorpusLinguistics* 9–2, pp.95–224.
46. Utkā, A., Levāne-Petrova, K., Bielinskienė, A., Kovalevskaitė, J., Rimkutė, E., Vēvere, D. 2012. Lithuanian–Latvian–Lithuanian Parallel Corpus. *Human Language Technologies – The Baltic Perspective*. *Proceedings of the Fifth International Conference Baltic HLT*, pp.260–264.
47. Vaičėnonienė, J. 2011. Lithuanian Literature in English. A Corpus-Based Approach to the Translation of Author-Specific Neologisms. (Daktaro disertacija). Kaunas: VDU.
48. Vaičėnonienė, J., 2012. Neįprastų kolokacijų vertimas grožinės literatūros kūriniuose: tekstynų paremtas tyrimas. *Darbai ir dienos*. Kaunas: VDU, pp.117–136.
49. Vaičėnonienė, J., 2014. How Relevant is the Distinction between the Original and Translated Texts in General Corpora? *Human Language Technologies – The Baltic Perspective*. Amsterdam/ Berlin: IOS Press, pp.248–256.
50. Vanderauwera, R., 1985. Dutch Novels Translated into English. *The Transformation of a 'Minority' Literature*. Amsterdam: Rodopi.
51. Volungevičienė, S., 2013. Kolokacijų vertimo problemos: gretinamoji vokiečių-lietuvių kalbų studija. (Daktaro disertacija). Vilnius: VU.
52. Xiao, R., He, L., Yue, M., 2010. In Pursuit of the Third Code: Using the ZJU Corpus of Translational Chinese in Translation Studies. In: Xiao, R., (ed.), *Using Corpora in Contrastive and Translation Studies*. Newcastle: Cambridge Scholars, pp.182–214.
53. Xiao, R., 2010. *Using Corpora in Contrastive and Translation Studies*. Cambridge Scholars Publishing.
54. Zanettin, F., 2012. *Translation-Driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome Publishing, pp.47–48.
55. Zanettin, F., 2013. Corpus Methods for Descriptive Translation Studies. *Social and Behavioral Sciences*, 95, pp.20–32. <https://doi.org/10.1016/j.sbspro.2013.10.618>

Šaltiniai

1. *4th Using Corpora in Contrastive and Translation Studies Conference (UCCTS)*. Prieiga per internetą: <http://ucrel.lancs.ac.uk/uccts4/past-confs.php> [Žiūrėta 2016 m. liepos mėn.].
2. *COMPARA: Portuguese - English Parallel Translation Corpus*. Prieiga per internetą: <https://user.clarin.eu/resources/compara-portuguese-english-parallel-translation-corpus> [Žiūrėta 2016 m. liepos mėn.].
3. *Dabartinės lietuvių kalbos tekstynas*. Prieiga per internetą: <http://tekstynas.vdu.lt/tekstynas/> [Žiūrėta 2016 m. liepos mėn.].
4. *Data Management Plan*. Prieiga per internetą: <http://www.clarin-d.de/en/preparation/data-management-plan> [Žiūrėta 2016 m. liepos mėn.].
5. *Depositing Services*. Prieiga per internetą: <https://www.clarin.eu/content/depositing-services> [Žiūrėta 2016 m. liepos mėn.].
6. *English-Italian Translational Corpus (CEXI)*. Prieiga per internetą: http://universal.elra.info/product_info.php?cPath=37_39&products_id=1454 [Žiūrėta 2016 m. liepos mėn.].
7. *Informacija apie lygiagretųjį tekstyną*. Prieiga per internetą: <http://tekstynas.vdu.lt/page.xhtml?id=info-about-parallel-corpus> [Žiūrėta 2016 m. liepos mėn.].
8. *International Association for Translation and Cultural Studies (IATIS)*. Prieiga per internetą: <http://www.iatis.org/index.php/iatis-belo-horizonte-conference> [Žiūrėta 2016 m. liepos mėn.].
9. *LILA lygiagretusis tekstynas*. Prieiga per internetą: <http://tekstynas.vdu.lt/page.xhtml?id=parallelLILA> [Žiūrėta 2016 m. liepos mėn.].
10. *Linguistic Properties of Translations. A Corpus-Based Investigation for the Language Pair English-German (CROCO)*. Prieiga per internetą: http://fr46.uni-saarland.de/croco/index_en.html [Žiūrėta 2016 m. liepos mėn.].
11. *Lygiagretusis tekstynas*. Prieiga per internetą: <http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus> [Žiūrėta 2016 m. liepos mėn.].
12. *Non-English, Parallel & Multilingual Corpora*. Prieiga per internetą: <https://www.uow.edu.au/~dlee/corpora2.htm> [Žiūrėta 2016 m. liepos mėn.].
13. *Oslo Multilingual Corpus (OMC)*. Prieiga per internetą: <https://www.hf.uio.no/ilos/english/services/omc/> [Žiūrėta 2016 m. liepos mėn.].
14. *The English-Norwegian Parallel Corpus (ENPC)*. Prieiga per internetą: <http://www.hf.uio.no/ilos/english/services/omc/enpc/> [Žiūrėta 2016 m. liepos mėn.].
15. *The ZJU Corpus of Translational Chinese (ZCTC)*. Prieiga per internetą: <http://www.lancaster.ac.uk/fass/projects/corpus/ZCTC/> [Žiūrėta 2016 m. liepos mėn.].
16. *Translational English Corpus (TEC)*. Prieiga per internetą: <https://www2.fgw.vu.nl/resources/corpora/detail.php?idnr=46> [Žiūrėta 2016 m. liepos mėn.].

Summary

Jurgita Vaičenonienė, Jolanta Kovalevskaitė, Teresė Ringailienė. Corpus-Based Translation Studies: Research and Resources

In Corpus Based Translation Studies, it is claimed that there are certain features common to all translated texts which make them different from the originals. The majority of investigations on the features of translations deal with the dominant languages, such as English (Baker, 1993, 1995, 1996; Laviosa, 1997, 1998), German (Kenny, 2001) or Chinese (Xiao et al., 2010). However, it remains unclear whether these features are common to the translations of other typologically different or minor languages. Despite the prevalence of translations, there has been a lack of language resources representative of translated Lithuanian. The aim of this article is to overview major research and resources in Corpus Based Translation Studies and to present a compiled comparable corpus of original and translated Lithuanian. The 4 million word corpus includes fiction and popular science texts, and consists of four sub-corpora (original and translated fiction; original and translated popular science litera-

ture). It is believed that the created comparable corpus of original and translated Lithuanian will contribute to further research on translations and have a practical value for trainee and professional translators, lecturers and language editors, who will be able to use the publicly available corpus for a comparable search of the language use in original and translated Lithuanian.

Jurgita Vaičenonienė

Humanitarinių mokslų daktarė, Vytauto Didžiojo universiteto Užsienio kalbų, literatūros ir vertimo studijų katedros lektorė.

Mokslinės veiklos sritys

Vertimo studijos, vertimo pedagogika, tekstynų lingvistika, diskurso analizė.

Adresas

Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Užsienio kalbų, literatūros ir vertimo studijų katedra, V. Putvinskio g. 23-320, LT-44243 Kaunas.

El. paštas:

jurgita.vaicenoniene@vdu.lt

Jolanta Kovalevskaitė

Humanitarinių mokslų daktarė, Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro mokslo darbuotoja, Užsienio kalbų, literatūros ir vertimo studijų katedros lektorė.

Mokslinės veiklos sritys

Tekstynų lingvistika, frazeologija, leksikologija ir leksikografija, terminologija.

Adresas

Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Kompiuterinės lingvistikos centras, V. Putvinskio g. 23-216, LT-44243 Kaunas.

El. paštas:

jolanta.kovalevskaite@vdu.lt

Teresė Ringailienė

Humanitarinių mokslų daktarė, Vytauto Didžiojo universiteto Užsienio kalbų, literatūros ir vertimo studijų katedros lektorė.

Mokslinės veiklos sritys

Multimodalioji diskurso analizė, mokslo populiarinimo ir akademinis diskursas, tekstynų lingvistika, vertimo studijos.

Adresas

Vytauto Didžiojo universitetas, Humanitarinių mokslų fakultetas, Užsienio kalbų, literatūros ir vertimo studijų katedra, V. Putvinskio g. 23-320, LT-44243 Kaunas.

El. paštas:

terese.ringailiene@vdu.lt

Apie autores