

Palyginamasis originalios ir vertimų lietuvių kalbos tekstynas

**Jurgita Vaičenonienė, Jolanta Kovalevskaitė,
Teresė Ringailienė**

VDU Kompiuterinės lingvistikos centras
Užsienio kalbų, literatūros ir vertimo studijų katedra

Vertimų kalbos tyrimų aktualumas

- Vertimo studijose jau kurį laiką yra teigiama, kad versti ir originalūs tos pačios kalbos tekstai gali būti suprantami kaip skirtingi registrai su jiem būdingomis ypatybėmis.
- Šis požiūris ypač aktualus mažiau paplitusioms kalboms, kurios yra veikiamos dominuojančių kalbų.

2014 m. Lietuvos spaudos statistika:

- iš anglų kalbos verstos knygos sudarė 63% vertimų;
 - verstų knygų tiražas vidutiniškai dvigubai didesnis (1,9 tūkst. egz.) nei originalių (1,0 tūkst. egz.);
 - didžioji dalis vertimų yra grožinės literatūros kūriniai (60%).
-
- Lietuviškuose naujienų portaluose ir informaciniuose, mokslo populiarinimo bei pramoginiuose periodikos leidiniuose (pvz., *A-ZET*, *National Geographic Lietuva*, *Cosmopolitan*, *Iliustruotas mokslas*, ir kt.) nemažai pateikiamos informacijos yra verčiama iš įvairių užsienio šaltinių.
-
- Pramoginė grožinė ir įvairių sričių mokslo populiarinimo literatūra – daugiausiai skaitoma.

Vertimų kalbos tyrimų aktualumas

- Tekstynais paremti vertimų kalbos tyrimai (angl. *Corpus Based Translation Studies*) yra viena ryškiausių kryptių moderniose vertimo studijose.
- Daugiausia tyrimų – didžiųjų kalbų, todėl skatinami kuo įvairesnių kalbų porų ir kryptių tyrimai.

Vertimų kalbos tyrimai (plg. Laviosa 1998)

- Lyginant anglų vertimų kalbos tekstyno (*Translational English Corpus*) ir Britų nacionalinio tekstyno (*British National Corpus*) duomenis nustatyta, kad angliškuose vertimuose paprastinimą rodo mažesnė leksinė įvairovė ir tankumas, siauresnė informacinė apimtis:
 - santykis tarp visų ir skirtingų žodžių mažesnis nei originaliuose tekstuose;
 - aukšto dažnumo žodžių daugiau lyginant su retai pasitaikančių žodžių sąrašu;
 - dažniausi žodžiai pasikartoja dažniau nei originaluose;
 - prasminių žodžių mažiau nei tarnybinių.

Tekstynai vertimų kalbai tirti

- *Translational English Corpus.*
<https://www2.fgw.vu.nl/resources/corpora/detail.php?idnr=46>
- *The English-Norwegian Parallel Corpus ENPC.*
<http://www.hf.uio.no/ilos/english/services/omc/enpc/>
- *Oslo Multilingual Corpus (OMC).* <https://www.hf.uio.no/ilos/english/services/omc/>
- *CEXI (English Italian Translational Corpus).*
http://universal.elra.info/product_info.php?cPath=37_39&products_id=1454
- *Linguistic Properties of Translations. A Corpus-Based Investigation for the Language Pair English-German (CROCO).* http://fr46.uni-saarland.de/croco/index_en.html
- *The ZJU Corpus of Translational Chinese (ZCTC).*
<http://www.lancaster.ac.uk/fass/projects/corpus/ZCTC/>
- *COMPARA : Portuguese - English Parallel Translation Corpus.*
<https://user.clarin.eu/resources/compara-portuguese-english-parallel-translation-corpus>

Tikslas

- Pristatyti naują šaltinį vertimų kalbai tirti – kuriamą palyginamąjį lietuvių kalbos tekstyną:

*Palyginamasis **originalios ir vertimų lietuvių**
kalbos **tekstynas (ORVELIT)***

Esami šaltiniai vertimams tirti: iš projektų sukaupiti resursai

- *Lygiagretusis tekstynas* (dvikryptis anglų-lietuvių; dvikryptis čekų-lietuvių). en->lt sudaro 2 024 999 žodžių; lt->en sudaro 61 154 žodžiai

<http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus>

- *LILA lygiagretusis tekstynas*: 9 mln. žodžių dvikryptis lietuvių-latvių-lietuvių tekstynas.

<http://tekstynas.vdu.lt/page.xhtml?id=parallelLILA>

Disertacijų tyrimuose sukaupti resursai

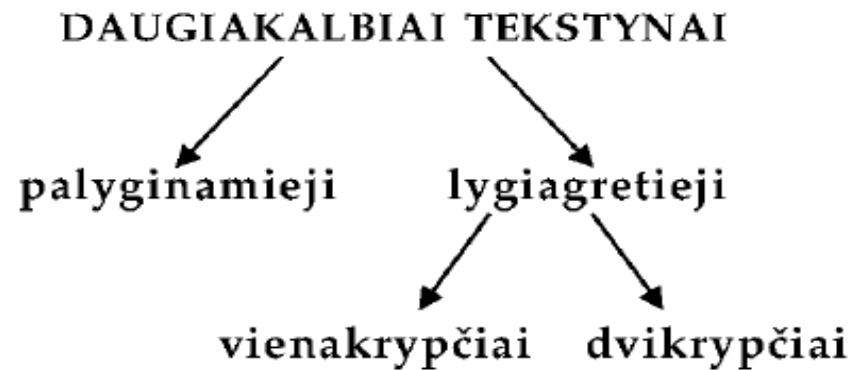
- **Anglų ir lietuvių** kalbos:

- lygiagretusis grožinės lietuvių literatūros vertimų į anglų kalbą tekstynas (1,690164 mln., žodžių, 759841 lygiagrečių sakinių) (Vaičenonienė 2011);
- dvikalbis dvikryptis lygiagretusis anglų-lietuvių kalbų tekstynas (1,572498 mln. žodžių), sudarytas iš originalių grožinės literatūros tekstų ir jų vertimų (Šolienė 2013).

- **Vokiečių ir lietuvių** kalbos:

- lygiagretusis vokiečių-lietuvių kalbų tekstų ir jų vertimų tekstynas (1 mln. žodžių), susidedantis iš grožinės literatūros kūrinių bei jų vertimų (Kovalevskaitė 2012);
- lygiagretusis vokiečių-lietuvių kalbų tekstų ir jų vertimų tekstynas (1 mln. žodžių), susidedantis iš įvairių kanceliarinių, administracinių, grožinės literatūros ir populiariojo mokslo kūrinių bei jų vertimų (Volungevičienė 2013).

Daugiakalbiai tekstynai (Rimkutė, Kovalevskaitė, Daudaravičius 2006)



1 pav. Daugiakalbių tekstynų klasifikacija

Palyginamieji tekstynai (Rimkutė, Kovalevskaitė, Daudaravičius 2006)

- **Vienakalbiai** palyginamieji tekstynai

(angl. *monolingual comparable corpus*) –

sudaryti iš panašios tematikos verstų ir neverstų tos pačios kalbos tekstų, pvz. pristatomas tekstynas ORVELIT.

- **Daugiakalbiai** palyginamieji tekstynai

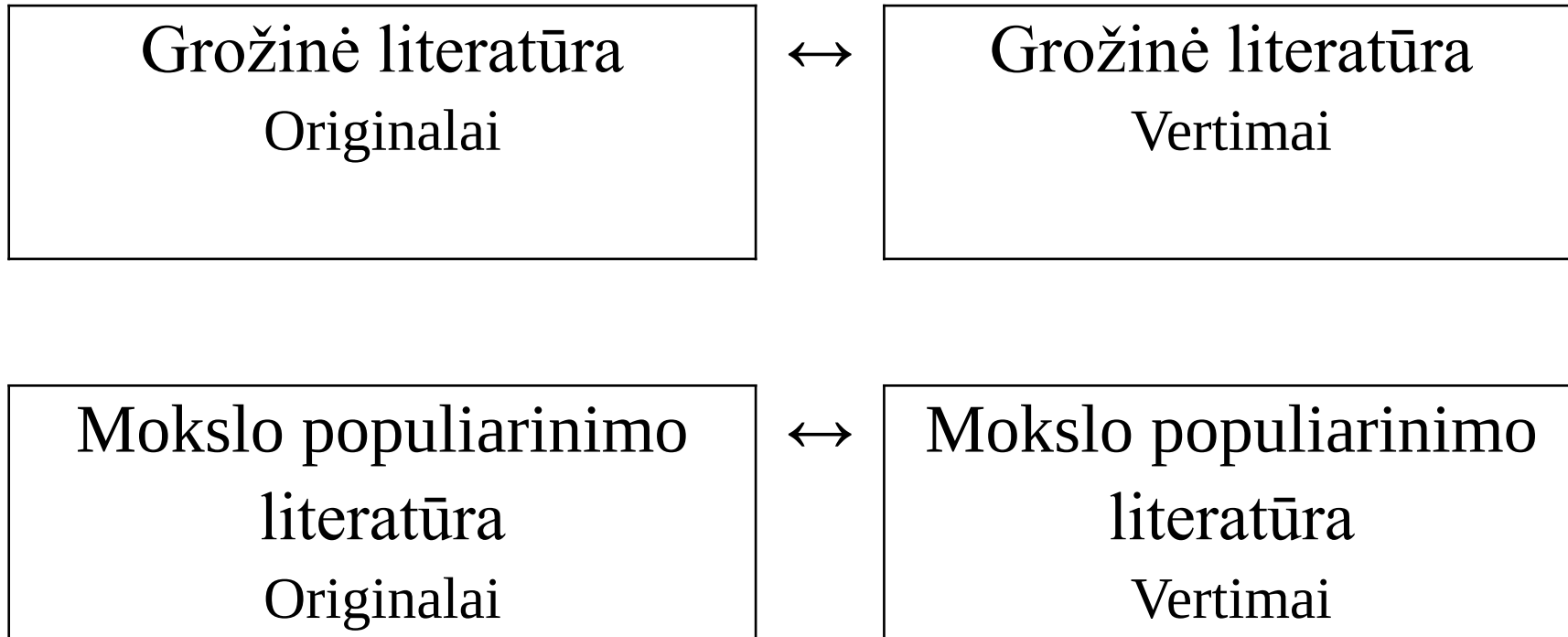
(angl. *multilingual comparable corpus*) –

sudaryti iš panašios tematikos verstų ir neverstų skirtingų kalbų tekstų (pvz., prancūzų ir lietuvių kalbų palyginamasis tekstynas politiniam diskursui tirti (Leonavičienė 2005)).

ORVELIT SUDARYMO TIKSLAS

- Reprezentuoti vertimų kalbą, pasiūlant tyrėjams resursą leksinėms, gramatinėms, sintaksinėms vertimų kalbos ypatybėms tirti.
- Kol kas panaudojant ORVELIT galima būtų tirti minėtų ypatybių pasiskirstymą grožinėje literatūroje ir mokslo populiarinimo tekstuose.
- Numatyta tekstyno apimtis – 4 mln. žodžių; originaliosios ir vertimų kalbos grožinės literatūros ir mokslo populiarinimo patekstynių – po 1 mln. žodžių.

ORVELIT patekstiniai



Grožinės literatūros dalis

Kriterijus	Grožinės literatūros vertimų patekstynis	Originalios grožinės literatūros patekstynis
<i>Tipas</i>	Specializuotas, lyginamasis, vienkaltis, vienkryptis.	Specializuotas, lyginamasis, vienkaltis, vienkryptis.
<i>Reprezentatyvumas</i>	Šiuolaikinės anglų kalbos grožinės literatūros vertimai į lietuvių kalbą.	Šiuolaikinė lietuvių literatūra.
<i>Populiacija</i>	Šiuolaikinės anglų kalbos grožinės literatūros vertimai į lietuvių kalbą.	Šiuolaikinė lietuvių literatūra.
<i>Imties rėmai (sampling frame)</i>	<i>Dabartinės lietuvių kalbos tekstyno</i> grožinės literatūros dalis; <i>Lygiagrečiojo</i> tekstyno anglišku kūrinių vertimų dalis.	<i>Dabartinės lietuvių kalbos tekstyno</i> grožinės literatūros dalies kūriniai, rašyti lietuvių autorių.
<i>Imties vienetas (sampling unit)</i>	Knyga tekstinio failo formatu.	Knyga tekstinio failo formatu.
<i>Žodžių skaičius</i>	1033520	1009246

Grožinės literatūros dalis

<i>Tekstų skaičius</i>	17 visateksčių kūrinių	19 visateksčių kūrinių
<i>Žanrai</i>	Rimtoji ir populiarioji literatūra	Rimtoji ir populiarioji literatūra
<i>Autoriai</i>	17 autorių	19 autorių
<i>Autoriai pagal lytį</i>	7 moterys, 10 vyrų	10 moterų, 9 vyrai
<i>Vertėjai</i>	15 vertėjų	-
<i>Leidyklos</i>	10 leidyklų	5 leidyklos
<i>Chronologiniai vertimų rėmai</i>	Vertimai po Nepriklausomybės atkūrimo 1990 m. (1997–2011 m.)	-
<i>Chronologiniai kūrinių rėmai</i>	Šiuolaikinė literatūra (XX a. antroji pusė – XXI a.)	Šiuolaikinė literatūra (XX a. antroji pusė – XXI a.)

Mokslo populiarinimo samprata

- Mokslo populiarinimas gali būti suprantamas keleriopai: kaip visuomenės susidomėjimas mokslu, kartais tiesiog kalbant apie visuomenės supratimą apie mokslą, kalbant apie mokslinę kultūrą ar mokslinį raštingumą (Burns ir kt., 2003)
- Bendriausias mokslo populiarinimo supratimas – įvairios veiklos bei žanrai, kurie specializuotą mokslą transformuoja į nespecialistams suprantamą ir taip siekia juo sudominti (Calsamiglia ir van Dijk, 2004).

Mokslo populiarinimo dalis

Kriterijus	Vertimai	Originalai
<i>Tipas</i>	Specializuotas, lyginamasis, vienkaltis, vienkryptis.	Specializuotas, lyginamasis, vienkaltis, vienkryptis.
<i>Reprezentatyvumas</i>	Mokslo populiarinimo tekstai – vertimai iš anglų kalbos į lietuvių.	Mokslo populiarinimo tekstai lietuvių kalba.
<i>Populiacija</i>	Mokslo populiarinimo knygos – vertimai iš anglų kalbos į lietuvių.	Originalios, lietuvių kalba rašytos mokslo populiarinimo knygos.
<i>Imties rėmai (sampling frame)</i>	<i>Dabartinės lietuvių kalbos tekstyno</i> negrožinės literatūros dalis; <i>Lygiagrečiojo</i> tekstyno angliškų kūrinių vertimų dalis.	<i>Dabartinės lietuvių kalbos tekstyno</i> negrožinės literatūros knygos, rašytos lietuvių autorių.
<i>Imties vienetas (sampling unit)</i>	Knyga tekstinio failo formatu.	Knyga tekstinio failo formatu.
<i>Žodžių skaičius</i>	877473	1,222802

Mokslo populiarinimo dalis

<i>Tekstų skaičius ir apimtis</i>	16 visateksčių kūrinių	26 visateksčiai kūrinių
<i>Sritys</i>	Istorijos, psichologijos, edukologijos, biologijos, verslo, mokslo populiarinimas	Istorijos, verslo, psichologijos, teisės, politikos, informacinių technologijų mokslo populiarinimas
<i>Autoriai</i>	16 autorių	25 autoriai
<i>Autoriai pagal lytį</i>	6 moterys, 9 vyrai	11 moterų, 11 vyrų
<i>Vertėjai</i>	16 vertėjų	-
<i>Leidyklos</i>	7 leidyklos	11 leidyklų
<i>Chronologiniai vertimų rėmai</i>	Vertimai atlikti 2005-2012 metais	-
<i>Chronologiniai kūrinių rėmai</i>	-	2005-2012

Kuriant ORVELIT kaip lingvistinį išteklių vadovaujamasi duomenų tvarkybos žingsniais (Trippel ir Zinn (2015: 71)):

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 (ištekliaus kūrimo ir saugojimo biudžetas, darbų paskirstymas, etc.)

Tolesni ORVELIT rengimo darbai:

- anotavimas;
- tekstyno sąsaja;
- prieiga per CLARIN-LT saugyklą su akademinio pobūdžio (ACA) licencija.

Analizės įrankiai, skirti atlikti paiešką tekstyne

- Planuojama ORVELIT anotuoti morfologiškai (vėliau – ir sintaksiškai).
- Programinė įranga, kurią planuojama naudoti ORVELIT analizei, yra Berlyno Humboltų universitete sukurtas ir viešai prieinamas tekstyno analizės įrankių paketas ANNIS (<http://corpus-tools.org/annis/aql.html>).

Autorinių teisių klausimas

Kaupiant tekstus ORVELIT remiamasi *Teisinėje CLARIN-LT veiklos studijoje* (2015) pateiktu išaiškinimu dėl tekstynams renkamų tekstų autorių teisių:

[...] 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: 23).

Platinimo ir saugojimo klausimas

- Siejamas su CLARIN-LT – Bendrosios kalbos išteklių ir technologijų infrastruktūros – teikiamomis galimybėmis.
- 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.

Prieiga prie tekstyno

- Prieiga prie tekstyno bus galima per CLARIN-LT saugyklą su akademinio pobūdžio (ACA) licencija, kuri apibrėžia naudotojo tapatybės ir prieigos sąlygas, bendrąsias naudojimo ir platinimo sąlygas (*Teisinė CLARIN-LT veiklos studija*, 2015: 29-30).

ORVELIT panaudojimo galimybės

- vertėjus rengiantiems vertimo studijų ar filologijos dėstytojams;
- vertimo kalbą analizuojantiems tyrėjams;
- studentams – būsimiems vertėjams;
- praktikams (profesionaliems vertėjams, kalbos redaktoriams, vertimų kokybės vertintojams).

Literatūra

- Advokatų profesinė bendrija FORT. 2015. *Teisinė CLARIN-LT veiklos studija*. Vilnius.
- Baker, M. 1995. Corpora in Translation Studies. An Overview and Some Suggestions for Future Research. *Target* 7 (2), 223–243.
- Baker, M. 1996. Corpus-based Translation Studies – the Challenges that Lie ahead. Somers, H. *Terminology, LSP and Translation*. Philadelphia/ Amsterdam: John Benjamins, 175-86.
- Burns, T. W., O'Connor, D. J., and S. M. Stockmayer. 2003. Science Communication: A Contemporary Definition. *Public Understanding of Science* 12 (2), 183-202.
- Calsamiglia, H. and T. van Dijk. 2004. Popularization Discourse and Knowledge about the Genome. *Discourse and Society* 15 (4), 369-389.
- Laviosa, S. 1998. Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta* 43 (4), 557–570.
- Leonavičienė A. 2005: *Lietuvių ir prancūzų politinės spaudos tekstų heterogeniškumas*. Daktaro disertacija. Kaunas: VDU.
- Lietuvių skaitymo ypatumai*. 2015. *Veidas*. <http://www.veidas.lt/lietuviu-skaitymo-ypatumai>.
- Markevičienė, R., Tamulynienė, L. 2015. *Lietuvos spaudos statistika*. Vilnius: Lietuvos nacionalinė Martyno Mažvydo biblioteka.
- Rimkutė E., Kovalevskaitė J. Daudaravičius V. 2006: [*Daugiakalbių tekstynų naudojimas ir taikymas*](#). *Darbai ir Dienos* 45, 41–62.
- Šolienė, A. 2013. *Episteminio modalumo ekvivalentiškumo parametrai anglų ir lietuvių kalbose*. Daktaro disertacija. Vilnius: VU.
- Trippel, T., and Zinn, C. 2015. DMPTY – A Wizard for Generating Data Management Plans. *CLARIN 2015 Selected Papers. Linköping Electronic Conference Proceedings*, No. 123. <http://www.ep.liu.se/ecp/123/006/ecp15123006.pdf>
- Vaičenonienė, J. 2011. *Lithuanian Literature in English. A Corpus-Based Approach to the Translation of Author-Specific Neologisms*. Daktaro disertacija. Kaunas: VDU.
- Volungevičienė, S. 2013. Kolokacijų vertimo problemos: gretinamoji vokiečių–lietuvių kalbų studija. Daktaro disertacija. Vilnius: VU.