

# Tekstynai ir vertimai. Tekstynų naudojimas verčiant – vis dar akademinė prabanga?

Jonė Grigaliūnienė

Vilniaus universitetas  
Filologijos fakultetas  
Literatūros, kultūros ir vertimo tyrimų institutas  
[jone.grigaliuniene@flf.vu.lt](mailto:jone.grigaliuniene@flf.vu.lt)

**Santrauka.** Straipsnyje kalbama apie tekstynų lingvistikos įtaką kalbų studijoms, vertėjų rengimui, bandoma reflektuoti dabartinę situaciją, aptarti problemas ir iššūkius, su kuriais susiduriama. Straipsnyje taip pat bandoma atsakyti į Bowker dar 2004 metais iškeltą klausimą, ar tekstynų naudojimas verčiant yra akademinė prabanga, ar būtinybė (Bowker 2004). Atlikta apklausa rodo, kad šiuandieniniai vertėjai, vertėjų rengėjai ir studentai susipažinę su tekstynais, jais naudojami, tačiau pirmenybę vis dar teikia kitiems informacijos šaltiniams: internetui, įvairioms duomenų bazėms, enciklopedijoms ir vertimo atmintims. Straipsnyje kalbama apie tokio pasirinkimo priežastis ir teigiama, kad ateityje tekstynų išteklių naudojimas vertimo srityje priklausys nuo informuotumo apie tekstynų naudą rengiant vertėjus, nuo tekstynų išteklių, programų ir įrankių prieinamumo.

**Pagrindiniai žodžiai:** vertimas, tekstynai, apklausos, *Sketch Engine*.

## Corpora and Translation. Are Corpora Still an Academic Luxury?

**Abstract.** This paper aims to consider the impact corpora have made on language studies and to touch upon the interface between corpora use and translator training/practice. A small-scale survey conducted among the translation trainers/professionals and translation students, with the aim of finding out whether professional translators and students are aware of the existence of corpora and to what extent they use them in their work, revealed that both the trainers and the students are well aware of corpora, but they still prefer translation memory technology to using corpora when translating. They have also pointed out that they would be interested in a service which quickly provided domain-and-language specific corpora tailored to their needs and a tool for extracting terminology from a domain specific corpus. The paper presents a tool which is now widely available for academic institutions in Europe and which gives a chance to quickly and easily compile a specific corpus, extract keywords, provides concordances and gives a useful word sketch that could be of great help when translating. The paper concludes that corpora have yet to make an impact on translation studies and that this will depend on raising awareness of the usefulness of corpora for translation training and practice and the availability of corpora tools that could meet translator needs.

**Keywords:** translation, corpora, surveys, *Sketch Engine*.

## Įvadas

Tekstynai jau įsitvirtino kalbų studijų ir ypač svetimų kalbų mokymo srityse, nors tekstynų lingvistikos kelias ir nebuvo lengvas. Pirmieji kompiuteriniai tekstynai atsirado prieš penkis dešimtmečius, tada jų egzistavimas buvo labai sudėtingas ir miglotas. Praėjusio amžiaus šeštajame ir septintajame dešimtmečiuose kalbotyroje dominavo Chomskio mokykla. Šios mokyklos nuostatos buvo priešingos tekstynų lingvistikos pagrindiniams teiginiams, o pats Chomskis – labai kritiškai nusiteikęs tekstynų atžvilgiu. Tekstynų lingvistikos esmė – tekstai, empirinė medžiaga, o Chomskis teigė, kad lingvistą turi dominti kompetencijos dalykai, bet ne atliktis. To meto tekstynų lingvistikos atstovai dirbo labai sunkiomis sąlygomis. Vienas iš pirmojo kompiuterinio tekstyno (*the Brown Corpus*) sudarytojų Francis (1982: 7–8), paklausus, ką veikia, atsakė, kad gavo finansavimą tekstynui kaupiti. Pašnekovas pareiškė, kad tai visiškas valstybės pinigų švaistymas: jis, kaip gimtakalbis, per kelias minutes gali pateikti tiek pavyzdžių, kiek negalys joks tekstynas. Kitas to paties tekstyno kūrėjas Kučera (1991: 402–403) atsiminimuose rašė savo archyve saugą iš to meto žymaus lingvisto gautą laišką, kuriame pastarasis, prasilenkdamas su geru skoniu, perfrazavo gerai žinomą Goeringo posakį: „Kai išgirstu žodį kompiuteris, mano ranka pati siekia ginklo“ (*Whenever I hear the word computer, I reach for my gun*).

Tuo metu daugybė kalbininkų domėjimąsi skaičiavimo mašinomis laikė kalbotyros išdavyste. Vis dėlto visi puikiai žinome, kad istorija dažnai būna negailestinga tokiems kategoriškiems, teisuoliškiems teiginiams. Šiandien jau niekas neabejoja kompiuterių nauda, neabejoja ir tekstynų lingvistikos svarba, ypač dirbtinio intelekto kontekste. Neabejojo ir truputį anksčiau, nes dabar kalbame apie visai naujos kartos tekstynus. Pirmieji kompiuteriniai tekstynai, kaip antai minėtasis *Brown* tekstynas, pasirodė septintajame dešimtmetyje ir apėmė 1 milijoną žodžių (tuomet tai atrodė didžiulė apimtis). Tarp antrosios kartos tekstynų galima paminėti *Cobuild* tekstyną, kuris, iš pradžių sudarytas iš 7 ar 8 milijonų žodžių, vėliau išaugo į vadinamąjį tęstinį (*monitor*) tekstyną. O šiais laikais jau kalbama apie pasaulinį tinklą kaip didžiulį tekstyną, kurio apimties net neįmanoma nusakyti. Prie to dar sugrįšime, o dabar pakalbėkime apie tai, kam gi reikalingi tie tekstynai ir ką jie gali duoti tyrėjui.

## Kam reikalingi tekstynai

Tekstynai – gan didelės apimties tekstų rinkiniai, sudaryti pagal tam tikrus kriterijus taip, kad kuo geriau atspindėtų kalbą ar jos atmainą (McEnery 1996: 177). Tai dokumentuotos autentiškos kalbos rinkiniai, objektyvus kalbos vartosenos šaltinis, suteikiantis patikimumo, patikrinamumo, įrodomumo galimybę ir atskleidžiantis daug

naujų, netikėtų faktų apie kalbos vartoseną. Sinclairis (1991: 100) yra pasakęs, kad „kalba atrodo kitaip, kai mums tampa prieinama didžiulė jos apimtis ir iš karto“ (*Language looks rather different when you look at a lot of it at once*).

Technologijos lemia kitoki kalbotyros objekto supratimą, nes leidžia pamatyti kalbos modelius, kurie šiaip nėra matomi (Stubbs 2009: 2). (*Technology leads to a radically revised perception of the object of study for linguistics, because it becomes possible to observe patterns of language use which are otherwise invisible*).

Ir iš tiesų Sinclairis, kai aštuntajame dešimtmetyje Birmingamo universitete kartu su leidykla *Collins Publishers* ėmėsi *Cobuild* projekto, manė, kad kompiuteris tame projekte atliks tik pagalbinės priemonės vaidmenį: padės kaupti ir saugoti tekstus, atlikti kai kuriuos skaičiavimus. Vis dėlto labai greitai jis pamatė, kad sukaupti duomenys reikalauja peržiūrėti daugybę iki tol laikytų nepajudinamomis leksikografijos nuostatų, kad tekstynas verčia kurti visai naują leksikografijos teoriją ir kitokią leksikografijos kalbą.

## Tekstynai ir leksika

Tekstynų lingvistui apskritai sunku suprasti, kaip galima ką nors išmokti, jei nesinaudojama tekstynais. Leksiką galima studijuoti paprasčiausiai tyrinėjant atskirus leksinius vienetus ir jų vartoseną. Kokią informaciją gali suteikti tekstynai? Leksikai tyrinėti labiausiai tinka dideli tekstynai (tie, kurių dydis – ne mažiau kaip 50 milijonų žodžių), nes mažuose tekstynuose retesnės leksikos iš viso nerasime. Tai susiję su žodžių dažnumu kalboje. Pirmasis neįprastas dažniausių kalbos žodžių ypatybes aprašė amerikiečių lingvistas Zipfas (1935). Jis pažymėjo, kad dažniausios žodžių formos pasižymi neįprastai dideliu dažnumu ir užima neproporcingai didelę teksto dalį. Britų mokslininkė Moon pateikė tokią lemų pasiskirstymo lentelę:

**1 lentelė.** Lemų pasiskirstymas anglų kalbos banko tekстыne (*BoE*)

Rangas	Procentai tekстыne	Apytikris vartosenos skaičius
Dažniausi 10	23,5	Daugiau kaip 8000 kartų milijone
Dažniausi 100	44,9	Daugiau kaip 1000 kartų milijone
Dažniausi 1000	68,5	Daugiau kaip 100 kartų milijone
Dažniausi 10 000	88,9	Daugiau kaip 5 kartai milijone
Dažniausi 25 000	92,5	Daugiau kaip 1 kartas milijone
Dažniausi 100 000	94,5	Daugiau kaip 1 kartas 20 milijonų

Šaltinis: Moon 2010: 198

Dažniausi anglų kalbos žodžiai yra pagalbiniai, gramatiniai žodžiai. Anglų kalbos banke dešimt dažniausių žodžių yra šie: *the, be, of, and, a, in, to* (bendraties dalelytė), *have, to* (prielinksnis), *it*. Tačiau įdomesni vadinamieji prasminiai žodžiai, patenkantys į dažniausių žodžių šimtuką: daiktavardžiai: *year, time, person (people), day, man, way*, veiksmazodžiai *say, go, make, get, take, know, see, come, think, give* ir būdvardžiai *new, good*. Štai kodėl leksikos tyrimams reikia didžiulių tekstynų. Bendrieji tekstynai gali suteikti informacijos ir apie žodžių darybą: derivatus ar sudurtinius žodžius. Ne tekstynais pagrįsti tyrimai dažnai būna nedidelės apimties, deskriptyvūs, pasikartojantys, smulkūs, nesisteminiai.

Tekstynai padeda atskleisti leksinės raiškos dėsningumus (*lexical patterns*), apie kuriuos mes net nenutuokėme. Apskritai tekstynais pagrįstos ar tekstynų inspiruotos studijos parodė, kad leksika nėra, kaip teigė Chomskis, chaotiška ir bestruktūre (*unordered list of lexical formatives*). Priešingai, Sinclairis savo leksikografiniais ir kitais darbais įrodė, kad leksika vaidina daug svarbesnį, net centrinį vaidmenį kalbos struktūroje ir kalbinės reikšmės kūrimo procese.

Kokių netikėtumų atskleidė tekstynų studijos? Tekstynais pagrįstos studijos parodė, kad dauguma žodžių sudaro gana pastovius modelius. Stubbs (2001: 81) pateikė pavyzdžių su prieviksiu *brightly*, kurį jis tyrė pasiremdamas *Cobuild* tekstynu. *Cobuild* tekстыne šis žodis sutinkamas 1467 kartus, iš jų net 26 proc. atvejų – kartu su būdvardžiu *coloured: brightly* 1467 < *coloured* 26 proc. Toks dažnas junglumas nėra išskirtinis dalykas. Stubbs pateikia ir daugiau panašaus junglumo pavyzdžių: *Cobuild* tekстыne žodis *calorie* sutinkamas 846 kartus, iš jų 29 proc. atvejų su būdvardžiu *low*; būdvardis *classical* – 5471 kartą, iš jų 22 proc. visų vartosenos atvejų jungiamas su žodžiu *music; profile* – 5584 kartus, iš jų 28 proc. atvejų jungiamas su žodžiu *high*; žodis *shuttle* iš 3453 pavartojimo atvejų 33 proc. sutinkamas kartu su žodžiu *space*. Jei skaičiuosime ir įvairias žodžių formas, skaičiai taps dar įspūdingesni: *cheering* – 1226 atvejai, iš jų su žodžiu *crowd* – 13 proc., su *crowds* – 6 proc., iš viso 19 proc. visų pavartojimo atvejų; *resemblance* – 1085 atvejai, iš jų 18 proc. su žodžiu *bears*, 11 proc. – su *bear*, 11 proc. – su *bore*, 4 proc. – su *bearing*, iš viso 44 proc. pavartojimo atvejų. Jei imsimė skaičiuoti dar ir sinonimus, hiponimus ir hiperonimus, skaičiai išaugs dar labiau: *Cobuild* tekстыne žodis *breakaway* sutinkamas 1379 kartus, iš jų su žodžiu *republic(s)* – 35 proc., žodžiais *group, faction, party* – 45 proc.; *cheering* šiame tekстыne pavartotas 1226 kartus, iš jų 19 proc. atvejų – su *crowd(s)*, 30 proc. – su *people, supporters, fans, audience; humanitarian* pavartotas 3933 kartus, 23 proc. – su žodžiu *aid*, 39 proc. – su žodžiais *relief, assistance, help*. Tokie duomenys liudija tam tikrą naują kalbinio tyrimo lygmenį, kurį kalbininkai pradeda atrasti ir tyrinėti tik dabar.

Tekstynai idealiai tinka kolokacijų tyrimams. Šie tyrimai irgi gali pateikti daug netikėtumų. Stubbs (2001: 97) nustatė, kad net 80 proc. kolokatų iš 38 000 veiksm-

žodžio ir daiktavardžio *cause* pavartojimo atvejų yra neigiami. Patys dažniausi kolokatai yra šie: *cause* < *problem(s)* 1806, *damage* 1519, *death(s)* 1109, *disease* 591, *concern* 598, *cancer* 572, *pain* 514, *trouble* 471 > (Stubbs 2001: 46). Tai reiškia, kad *cause* turi tendenciją būti siejamas su neigiamais reiškiniiais ir įvykiais. Ši tendencija dar nėra taip stipriai išreikšta, kad laikytume ją žodžio *cause* konotacija, tačiau tas reguliarumas yra tikrai stebinantis. Panaši situacija pasikartoja ir kitais atvejais. Pavyzdžiui, tyrėja Channell (2000: 53) tyrė frazę *roam the streets*. Ji nustatė, kad 113 *roam the streets* formų anglų kalbos banke (*Bank of English*) veiksniai yra daiktavardžiai *prostitutes, vagrant children, armed men, mobs, looters, youth gangs and neo-Nazis, vandals, wild dogs and bigots* (Channell 2000: 53), o veiksmi, su kuriais siejama frazė *roam the streets*, yra tokie: *searching for food, attacking people, stoning cars, randomly beating people, burning and looting and rioting*. Taigi, ši frazė paprastai siejama su veikla, kuri yra pavojinga, kelianti grėsmę, bauginanti. Tokių rezultatų negalima gauti pasikliaujant vien intuicija, introspekcija ar keliais pavyzdžiais iš žodyno. Tam reikia didelių tekstynų. Tokie faktai negaunami sąmoningai reflektuojant. Dar daugiau, juos labai sunku paaiškinti teoriškai: tai nei tiesioginė žodžio reikšmė, nei konotacija. Kyla klausimas, kas lemia tuos modelius, tas dažnai pasitaikančias kolokacijas? Tai jau teoriniai semantikos aspektai, kurių dar niekas išsamiau neaprašė ir kurie laukia išsamesnių tyrimų.

Tiesa, Sinclairis kalba apie kitokį reikšmės suvokimą tekstynų lingvistikoje. Jo teigimu, centrinis kalbos vienetas yra tarpinis darinys tarp žodžio ir sakinio, t. y. leksinis vienetas, dažniau nesutampantis nei sutampantis su žodžiu (Sinclair 1996: 114). Naujasis kompozicinis vienetas yra reikšmės, funkcijos ir struktūros kompleksas, nors tekstas ir galima vieneto raiškos įvairovė dažnai tą vienovę ir maskuoja (Marcinkevičienė 2000: 42). Sinclairis (1996) vartoja sąvoką „išplėstinis reikšmės vienetas“ (*extended unit of meaning*), kurią sudaro leksinis (žodžio kolokacijos), gramatinis (koligacijos), semantinis (bendras kolokatų laukas) ir pragmatinis (semantinė prozodija, vertinamieji konteksto elementai) lygmenys. Tognini-Bonelli (2000) pritaikė Sinclairio išplėstinio reikšmės vieneto sąvoką ir metodiką kalbų mokymo srityje ir parodė, kaip tekstynų inspiruota metodika padeda studentams patiems atrasti kalbą, jos dėsnius, formuluoti hipotezes remiantis konkrečių faktų analize, rasti atsakymus į išsikeltus klausimus ir pagrįsti tyrimus<sup>1</sup>.

<sup>1</sup> Sinclairio sukurta metodika taikoma ir Vilniaus universiteto vertimo bakalauro studijų tekstynų lingvistikos seminaruose. Studentai vykdo mini projektus: nedidelėse grupėse atlieka pasirinktos žodžių grupės analizę, taikydami Sinclairio išplėstinio reikšmės vieneto realizavimo analizės metodą.

## Tekstynai ir vertėjai

Nors šiuo metu tekstynai jau visiškai įsitvirtinę kalbotyroje ir niekas nebekelia klausimo dėl jų reikalingumo ir naudos, vertėjai jais naudotis neskuba. 2005–2007 metais vykdant projektą *MeLLANGE (Multilingual eLearning in Language Engineering)* buvo išplatinta anketa, kurios tikslas – susipažinti su profesionalių vertėjų požiūriu į tekstynų naudojimą verčiant ir nustatyti su vertimo technologijomis susijusius vertėjų poreikius, buvo gauti toliau aprašomi duomenys (Bernardini 2006: 19). Anketą užpildė 623 respondentai: 90,8 proc. profesionalių vertėjų iš Jungtinės Karalystės, Prancūzijos, Vokietijos ir Italijos ir 9,2 proc. studentų. 40,5 proc. respondentų pareiškė renkantys įvairią informacinę medžiagą, daugiau nei pusė jų (69,4 proc.) – tekstus elektronine forma, 46,9 proc. skaitė tuos tekstus, bet nesinaudojo jokia paieškos programa, tie, kurie naudojami paieškos sistema, tai darė teksto apdorojimo programos paieškos įrankiu (65,9 proc.), tik maža dalis (19 proc.) naudojosi konkordansu. Nors daugelis vertėjų nebuvo susipažinę su tekstynais, dauguma jų norėjo kuo daugiau apie juos sužinoti: 78,6 proc. respondentų domino priegios prie specialiųjų tekstynų teikimo paslauga, 77,9 proc. – terminų paieškos įrankis, 82,4 proc. rūpėjo tekstynų galimybės.

Panaši anketa 2019 m. buvo išplatinta tarp lietuvių vertėjų, vertimo studijų dėstytojų ir studentų (apklausoje dalyvavo Vilniaus universiteto Vertimo studijų katedros dėstytojai ir studentai). Atlikus apklausą paaiškėjo, kad profesionalūs vertėjai ir dėstytojai kaupia tekstus (83,3 proc.), daugiausia elektronine forma, kai kurie – ir elektronine, ir popierine forma. Visi respondentai teigė, kad yra susipažinę su tekstynais, dauguma (66,6 proc.) nurodė, jog naudoja juos versdami tekstus. 50 proc. respondentų naudojami tokiais specialiais įrankiais kaip *Linguee*, *Eur-Lex* ir pasitelkia teksto apdorojimo paieškos įrankius. 90 proc. respondentų naudojami vertimo atmintimi, konkordansą pasitelkia nedaugelis (apie 30 proc.). Visi respondentai teigė, kad juos domintų specialieji tekstynai ir įrankiai, padedantys nustatyti terminus specialiajame tekстыne.

Dauguma apklausoje dalyvavusių studentų taip pat teigė naudojančius tekstynus versdami (72,6 proc.). Tik 8 iš 33 apklaustų studentų pareiškė nesinaudojančius tekstynais: trys teigė abejojančius tekstynų nauda vertėjui, vienas parašė, kad paprasta *Google* paieška yra daug greitesnis ir paprastesnis būdas susirasti reikiamą informaciją, du nurodė, kad jiems trūksta informacijos apie tekstynų naudojimą verčiant.

Atlikta nedidelės imties apklausa rodo, kad situacija, palyginti su *MeLLANGE* apklausa, yra pasikeitusi<sup>2</sup>, tačiau vertėjai vis dar teikia pirmenybę ne tekstynams, o žodynams, duomenų bazėms ir panašioms ištekliams. Žinoma, priežasčių yra keletas.

<sup>2</sup> Naujausia tekstynų naudojimo vertime apžvalga, pateikta Verplaetse ir Lambrechts (2019), patvirtina šią mintį.

Pirma, labai mažai vertėjų apskritai yra susipažinę su tekstynais ir įvairiomis programomis kaip vertėjo įrankiais. Kalbant apie Lietuvos universitetus, tekstynų naudojimas specialioms vertimams nėra kaip nors sistemaiškai ir metodiškai įgyvendintas (nors apie tai jau nemažai kalbama vertėjų rengimo programose įvairiose institucijose). Antra, jei tokie ištekliai ir minimi vertėjų rengimo programose, jų dažnai nenaudoja net ir vertėjai profesionalai, nes tekstynų kaupimas ir naudojimas reikalauja nemažai laiko išteklių, o vertėjai dažnai yra spaudžiami laiko ir terminų. Kita vertus, dauguma tekstynais pagrįstų priemonių sukurtos turint omenyje ne vertėjus, o lingvistus tyrėjus. Nors ir egzistuoja tam tikros priemonės, kuriomis galėtų naudotis vertėjai, jos nebuvo sukurtos pagal jų poreikius. Kai kurios programos, kaip antai *MicroConcord* ir *MonoCon*, sukurtos atsižvelgiant į pedagoginius poreikius, kitos, pvz., *WordSmithTools*, *AntConc* ir *TextStat*, orientuotos į kalbos tyrėjus, kompiuterinės lingvistikos atstovus, leksikografus. Bernardini (2006: 21) teigia, jog tam, kad tekstynai būtų naudojami vertėjų, paieškos sistemos turi būti paprastesnės ir labiau pritaikytos vertėjams. Taigi, reikia atsižvelgti į specifinius vertėjų poreikius. O jie skiriasi nuo kitų akademinų grupių poreikių.

## Ką tekstynai gali duoti vertėjams

Tekstynai iš principo vertėjams nėra jokia naujiena. Vertėjai visuomet rinko ir renka „lygiagrečius tekstus“. Dar prieš tekstynų ir elektroninių tekstų atsiradimą vertėjai rinko, skaitė ir analizavo įvairius tekstus, ieškojo juose terminų atitikmenų, sudarinėjo žodynelius ir glosarijus. Dabar situacija pasikeitusi ir nebereikalauja iš vertėjų tokio ekstensyvaus tekstų skaitymo, tačiau lygiagrečių tekstų rinkimas ir jų paskaitinėjimas gali būti naudingas vertėjams, siekiantiems susipažinti su tam tikra sritimi arba tam tikrais žanrų ypatumais.

Vertimo studijose naudojami įvairūs tekstynai: palyginamieji (*comparable*), lygiagretieji (*parallel*), vienkaliai (*monolingual*), dvikaliai (*bilingual*), daugiakaliai (*multilingual*), vertimų tekstynai (*translation corpora*), vertimo atmintys (*translation memories*) ir pan.

Palyginamieji tekstynai (*comparable corpora*) sudaryti iš originalių kelių kalbų tekstų, į tekstyną įtrauktų pagal tuos pačius ar panašius atrankos kriterijus ir tekstyno komponavimo principus (Marcinkevičienė 2002). Palyginamieji tekstynai yra puikus šaltinis kelioms kalboms lyginti, tam tikros srities ar žanro ypatumams tirti, terminų ekvivalentams rinkti ir daugeliui kitų tikslų. Kitas vertėjų plačiau naudojamas tekstynų tipas – lygiagretusis tekstynas (*parallel corpus*). Lygiagretieji tekstynai – tai tokie tekstynai, į kuriuos įtraukti tekstai originalo kalba A ir jų vertimai į kalbą B (ar net kelias kalbas – C, D ir t. t.). Šis tekstynų tipas labai patogus siekiant nustatyti vertimo normas, plačiai naudojamas gretinamosiose studijose, taip pat tai labai patogus būdas pasimo-

kyti iš profesionalių vertėjų. Lygiagretūs tekstynai yra ideali priemonė vertėjams, tačiau jų yra mažai, o jiems sukaupti reikia daug laiko ir lėšų. Vertimo studijose populiarūs ir vertimo kalbos (*translation corpora*) tekstynai, nes dažnai teigiama, kad vertimai ne tik daro įtaką kalbai, bet ir yra savarankiška kalbos atmaina su jai būdingomis savybėmis: paprastinimu (*simplification*), eksplicitiškumu (*explicitation*), norminimu (*normalization*) ir neutralizavimu (*leveling out*) (cf. Baker 1996). Vertimo kalbos tekstynai gali padėti suvokti tarpžanrinius tekstų reikalavimus ir ypatumus, versto teksto ypatybes, įvertinti vertimų ir originalių tekstų skirtumus<sup>3</sup>. Tiek palyginamieji, tiek lygiagretieji bei vertimo tekstynai yra vertingi vertėjams, vertimo studijų dėstytojams, kalbų mokytojams ir studentams. Dažnai jie papildo vieni kitus. Lygiagretieji tekstynai – tai originalo ir vertimo kalbos tekstynai, todėl kartais yra vertinami įtariai, pvz., Teubert (1996: 247) teigia, kad „darbas su vertimais reiškia darbą su iškreiptu veidrodiniu originalo kalbos atvaizdu vertimo kalboje“. Taigi labai dažnai šalia lygiagrečiųjų tekstynų pasitelkiami ir palyginamieji tekstynai. Šie tekstynai gali papildyti ar net pakeisti kai kuriuos kalbinius šaltinius, tokius kaip žodynai ar enciklopedijos, kurių dažniausiai nepakanka profesionaliems vertėjams, kai reikia versti tekstus, susijusius su specialia sritimi, todėl vertėjai dažniausiai naudojami pasauliniu tinklu, kuriame, kaip teigė anketos respondentai, galima rasti beveik viską. Internetas, žinoma, gali padėti vertėjui, todėl šiandien galima kalbėti apie pasaulinį tinklą kaip apie milžiniškos apimties tekstyną (cf. Kilgariff 2001; Kilgariff, Grefenstette 2003; Gatto 2014). Dabar daug diskutuojama apie iš interneto šaltinių sugeneruoto tekstyno statusą. Tekstynas lingvistinėje literatūroje siejamas su tokiais sąvokomis kaip reprezentatyvumas, autentiškumas, balansas, tekstų ar pavyzdžių atranka, tekstyno dydis, tekstyno konstravimo principai arba kriterijai, tekstyno tikslai ir t. t. Iš interneto šaltinių sukauptas tekstynas neturi jokių galimybių atitikti tuos kriterijus, tačiau jei, kaip teigia Kilgariff ir Grefenstette (2003), kelsime ne ontologinį klausimą „Kas yra geras tekstynas?“, o daugiau praktinį – „Ar tas tekstynas tinka problemai X spręsti?“, tai internetas gali padėti spręsti tam tikras kalbines (ir ne tik) problemas. Viena iš priežasčių, lėmusių pasaulinio tinklo kaip tekstyno naudojimą, yra ta, kad pasaulinis tinklas – neginčijamas autentiškos kalbos šaltinis. Internete esantys tekstai yra autentiško bendravimo rezultatas, tačiau reikia atminti, kad internete autentiškumas nebūtinai siejamas su patikimumu. Autentiška gali būti ir netaisyklinga rašyba, ir gramatikos, leksikos, sintaksės klaidos, netinkama vartoseną. Pasaulinis tinklas negali reprezentuoti kalbos taip, kaip reprezentatyvumo sąvoka suprantama tekstynų lingvistikoje. Kita vertus, tame slypi ir pasaulinio tinklo

<sup>3</sup> Tokio tekstyno pavyzdys yra Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre kaupiamas tekstynas ORVELIT (žr. Vaičėnienė et al. 2017).



stiprybė: jis nėra sukonstruotas lingvistų, bet yra tiesioginio žmonių bendravimo rezultatas, todėl jei ne reprezentuoja, tai bent atspindi tai, kas vyksta tarptautinėje bendruomenėje realiaame laike (Henzinger, Lawrence 2004: 5186). Žinoma, negalima teigti, kad pasauliniame tinkle esantys tekstai reprezentuoja kalbą, tačiau iš interneto šaltinių sugeneruoti tekstynai savo apimtimi, įvairove išpūdingi ir tam tikra prasme tai yra atsvara reprezentatyvumui.

Su reprezentatyvumu susiję ir tekstyno apimties klausimai. Tekstynų lingvistai teigia (cf. McEnery, Wilson 1996), kad tekstynai yra baigtiniai (išskyrus tęstinius – *monitor corpora*), o tai svarbu statistiniams skaičiavimams, rezultatų lyginimui ir patikimumui. Iš internetinių šaltinių sugeneruotas tekstynas daro tokius skaičiavimus neįmanomus, taigi kyla abejonė dėl taip sudaryto tekstyno duomenų patikimumo, tekstyno turinio neapibrėžtumo, jo mokslinės vertės.

Per pastaruosius dvidešimt metų tekstynai ir tekstynų analizės programinė įranga ne kartą buvo minimi literatūroje kaip veiksmingas įrankis, galintis sukurti puikią kalbinę aplinką vertėjams (cf. Aston 2009; Bowker 1998; Fantinuoli 2013; Kübler 2011), gerinti vertimo kokybę (cf. Zanettin 2012; Varantola 2003). Be to, tekstynai gali pasiūlyti daug didesnę kalbos įvairovę nei žodynai, jie ugdo ir lavina kalbos jausmą, padeda plėsti kalbinės raiškos priemonių asortimentą (Teich 2003; Tognini-Bonelli 2001; Fantinuoli 2013). Skaitinėdami ir analizuodami konkordanso eilutes, vertėjai gali ne tik susipažinti su terminija, frazeologija ir idiomatika, bet ir įgyti faktinių žinių, nes net specializuodamiesi tam tikroje srityje dažniausiai verčiamą sritį išmano mažiau negu srities specialistai. 2 lentelėje pateikiamas informacijos, kurią gali suteikti tekstynai, pavyzdys.

**2 lentelė.** Ištrauka iš paieškos tekстыne, sudarytame iš pasaulinio tinklo

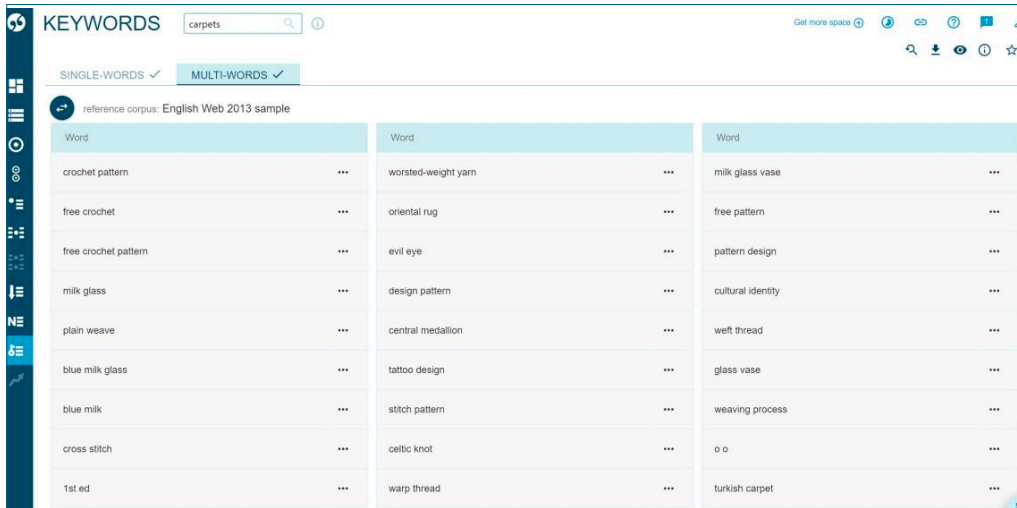
Originalas	Gentiniai <i>kilimai</i> yra nedideli, surištiper labai trumpą laiką – 3 ar 4 mėnesius.
Pavyzdžiai iš tekstyno (tekstynas sudarytas autorės {J.G. }naudojantis <i>Sketch engine</i> programa)	<p><i>The difference between a <b>kilim rug</b> and other types of <b>rugs</b> is that the design that is created on the kilim is made by interweaving the different colored wefts and warps, creating what is known as a flatweave.</i></p> <p><i>Turkish <b>rugs</b> and <b>kilims</b> are also frequently used.</i></p> <p><i>This motif is widely found in Kurdish and Turkish totemic <b>kilims</b> or flatweaves.</i></p> <p><i>Although the name of <b>kilim</b> is sometimes used loosely in the West to include all types of rugs...</i></p>

Informacija 2 lentelėje rodo, kad lietuviškas žodis *kilimas* gali turėti tokius angliškus atitikmenis: *rug*, *kilim*, *carpet*. Tekstynas, iš kurio paimta 2 lentelėje pateikta informacija, buvo (autorės – J. G.) sudarytas naudojantis programa *Sketch Engine*. Šią programą naudoja leksikografai, kalbų tyrėjai ir kalbų mokytojai. Jos skiriamasis bruožas yra „žodžio eskizas, aprašas“ (*word sketch*) – viename puslapyje telpanti žodžio gramatinės ir leksinės informacijos santrauka. Programa suteikia prieigą prie daugybės tekstynų penkiasdešimčia kalbų. Neseniai ji patobulinta, atsižvelgiant į vertėjų poreikius: papildyta automatiniais kolokacijų žodynais (*automatic collocation dictionaries*), lygiagrečiais tekstynais, dvikalbiais žodžių aprašais (*bilingual word sketches*), terminų ieškikliu (*term finder*). Dabartiniu metu (iki 2022 metų) programa nemokamai gali naudotis daugelis Europos Sąjungos akademinė institucijų (internetu adresas: *sketchengine.eu*). Programa *Sketch Engine* gali būti labai naudinga vertėjams: ji suteikia galimybę greitai ir lengvai sukaupti specializuotą tekstyną – tereikia įvesti keletą specialių žodžių (*seed words*) ir tekstai atsitiktine tvarka parenkami iš interneto (tekstų skaičių galima nustatyti). Tiesa, tekstynui sudaryti galima naudoti ir savo sukauptus tekstus. 2 lentelėje pateikti pavyzdžiai paimti iš tekstyno apie persiškus kilimus, sukaupto įvedus 10 specialių žodžių (per keliolika minučių programa sugeneravo daugiau negu 500 000 žodžių specialųjį tekstyną). Šis specialusis tekstynas suteikia įvairių galimybių: nustatyti raktažodžius ir terminus, kolokacijas, naudotis konkordanso programa, pateikia pasirinkto žodžio aprašą, arba eskizą (*sketch*) (žr. 3–5 lenteles).

3 lentelėje pateikiami raktažodžiai (*keywords*).

Word	Word	Word	Word	Word
motif ...	Navajo ...	medallion ...	ikat ...	Jamdani ...
weft ...	Patterns ...	paisley ...	Anatolia ...	Symbols ...
rug ...	Crochet ...	weaving ...	geometric ...	stitch ...
weaver ...	gul ...	Weaving ...	Baluchari ...	Motif ...
kilim ...	crochet ...	sc ...	Turkoman ...	carpet ...
Rugs ...	Balouch ...	Baluch ...	sl ...	Ch ...
Rug ...	Weave ...	twill ...	Persian ...	Lsc ...
weave ...	Kilim ...	saree ...	loom ...	heddle ...
Motifs ...	Anatolian ...	Turkmen ...	dc ...	Paracord ...

## 4 lentelėje pristatomi daugiažodžiai raktiniai junginiai.



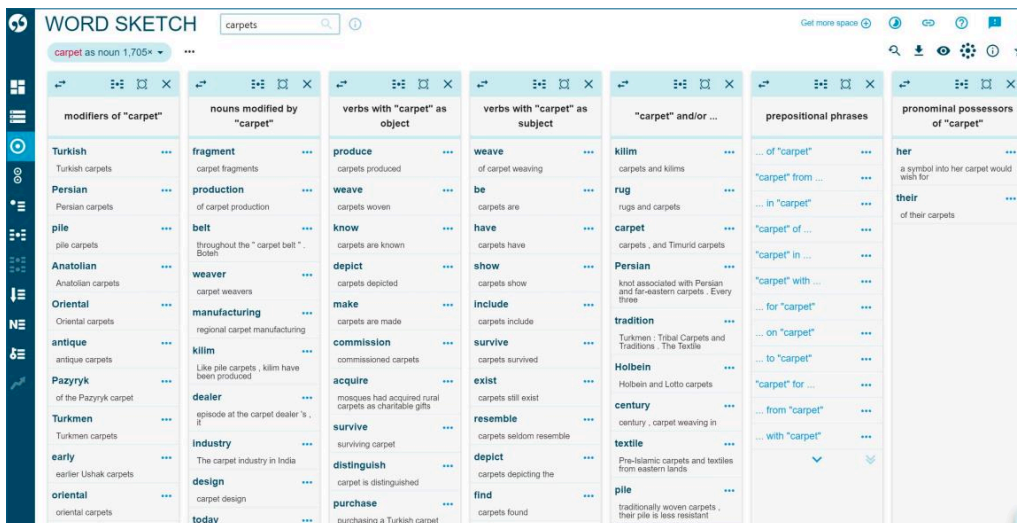
KEYWORDS carpets

SINGLE-WORDS ✓ MULTI-WORDS ✓

reference corpus: English Web 2013 sample

Word	Word	Word
crochet pattern	worsted-weight yarn	milk glass vase
free crochet	oriental rug	free pattern
free crochet pattern	evil eye	pattern design
milk glass	design pattern	cultural identity
plain weave	central medallion	weft thread
blue milk glass	tattoo design	glass vase
blue milk	stitch pattern	weaving process
cross stitch	celtic knot	o o
1st ed	warp thread	turkish carpet

5 lentelėje pateikiama įvairiapusė kalbinė informacija apie pasirinktą žodį (*word sketch*).



WORD SKETCH carpets

carpet as noun 1,705+

modifiers of "carpet"	nouns modified by "carpet"	verbs with "carpet" as object	verbs with "carpet" as subject	"carpet" and/or ...	prepositional phrases	pronominal possessors of "carpet"
<b>Turkish</b> ... Turkish carpets	<b>fragment</b> ... carpet fragments	<b>produce</b> ... carpets produced	<b>weave</b> ... carpets woven	<b>kilim</b> ... carpets and kilims	... of "carpet"	<b>her</b> ... a symbol into her carpet would wish for
<b>Persian</b> ... Persian carpets	<b>production</b> ... of carpet production	<b>weave</b> ... carpets woven	<b>be</b> ... carpets are	<b>rug</b> ... rugs and carpets	... in "carpet"	<b>their</b> ... of their carpets
<b>pile</b> ... pile carpets	<b>belt</b> ... throughout the "carpet belt". Bosnia	<b>know</b> ... carpets are known	<b>have</b> ... carpets have	<b>carpet</b> ... carpets ... and Timurid carpets	... "carpet" of ...	
<b>Anatolian</b> ... Anatolian carpets	<b>weaver</b> ... carpet weavers	<b>depict</b> ... carpets depicted	<b>show</b> ... carpets show	<b>Persian</b> ... knot associated with Persian and far-eastern carpets. Every three	... "carpet" with ...	
<b>Oriental</b> ... Oriental carpets	<b>manufacturing</b> ... regional carpet manufacturing	<b>make</b> ... carpets are made	<b>include</b> ... carpets include	<b>tradition</b> ... Turkmen - Tribal Carpets and Traditions. The Textile	... for "carpet"	
<b>antique</b> ... antique carpets	<b>kilim</b> ... Like pile carpets, kilims have been produced	<b>commission</b> ... commissioned carpets	<b>survive</b> ... carpets survived	<b>Holbein</b> ... Holbein and Lotto carpets	... on "carpet"	
<b>Pazyryk</b> ... of the Pazyryk carpet	<b>dealer</b> ... episode at the carpet dealer's ... it	<b>acquire</b> ... mosques had acquired rural carpets as charitable gifts	<b>exist</b> ... carpets still exist	<b>century</b> ... century ... carpet weaving in	... to "carpet"	
<b>Turkmen</b> ... Turkmen carpets	<b>industry</b> ... The carpet industry in India	<b>survive</b> ... surviving carpet	<b>resemble</b> ... carpets seldom resemble	<b>textile</b> ... Pre-Islamic carpets and textiles from eastern lands	... "carpet" for ...	
<b>early</b> ... earlier Ushak carpets	<b>design</b> ... carpet design	<b>distinguish</b> ... carpet is distinguished	<b>depict</b> ... carpets depicting the	<b>pile</b> ... traditionally woven carpets ... their pile is less resistant	... with "carpet"	
<b>oriental</b> ... oriental carpets	<b>today</b> ...	<b>purchase</b> ... purchasing a Turkish carpet	<b>find</b> ... carpets found			

5 lentelėje kompaktiškai pateikiama informacija apie žodį *carpet* ir jo vartoseną. Daugiau nei dešimtmetį tokia informacija buvo pagrindinis šaltinis leksikografams, žodynų sudarytojams, tačiau ji gali būti naudinga ir vertėjams profesionalams, būsimiems vertėjams, kiekvienam, kas domisi kalba ir jos vartoseną. Darbas su tekstiniais turi būti

įtraukiamas į vertėjų rengimo programas, studentus reikia supažindinti su tekstynų teikiama galimybėmis. Nors tai jau vyksta, vertėjams tekstynai dar nėra tapę būtinybe. Kad taip įvyktų, vertėjai turi būti supažindinami su tekstynų lingvistikos naujovėmis, o pačios naujovės turi būti labiau orientuotos į vertėjus, atitikti jų poreikius, programos turi būti lengvai prieinamos, turi būti paprasta jomis naudotis. Be to, būsimieji vertėjai turi būti ne tik supažindinami su naujomis programomis, bet ir turėti galimybių tuos įrankius išbandyti, patirti ir įvertinti jų naudą.

## Apibendrinimas

Tekstynai skinasi kelią į vertimo studijas, vertėjų rengimo programas ir profesionalių vertėjų praktiką. Apklausos rodo, kad situacija šiandien, palyginti su pirmųjų apklausų, atliktų 2005–2007 metais, duomenimis, yra pasikeitusi ir jau neabejojama tekstynų nauda verčiant. Kita vertus, tekstynai, neskaitant nedidelių išimčių – vertimo atminties bei internetinio teksto, kaip antai *Linguee*, dar nėra vertėjų atrasti. Galima paminėti keletą priežasčių. Viena iš priežasčių yra ta, kad tekstynui sukaupti reikia daug laiko, pastangų ir išmanymo. Kita priežastis – daugelis tekstynais pagrįstų priemonių sukurtos atsižvelgiant į kalbininkų, o ne vertėjų poreikius. Norint, kad vertėjai naudotųsi tekstynais ir programine įranga, paieškos sistemos turi būti paprastesnės ir orientuotos į vertėjus. Tokios programos tampa vis prieinamesnės. Viena tokių programų – *Sketch Engine*, suteikia vertėjams galimybę greitai sukaupti specialųjį tekstyną, turi nesudėtingą paieškos sistemą, terminų nustatymo funkciją, pateikia kolokatų sąrašus, leidžia naudotis konkordanso programa. Vis dėlto tekstynai tebėra akademinė prabanga, paprastai naudojami tik vertėjų rengimo programose, ir dar nėra tapę nei kasdienybe, nei būtinybe.

## Literatūra

- Aston, Guy. 2009. Forward. *Corpus use in translating: Corpus use for learning to translate and learning corpus use to translate*, edited by Alison Beeby, Patricia Rodriguez-Inés, and Pilar Sánchez-Gijón, IX-X. Amsterdam: John Benjamins.
- Baker, Mona. 1996. Corpus-based Translation Studies – the Challenges that Lie Ahead. *Terminology, LSP and Translation*, edited by Harold Somers. Philadelphia / Amsterdam: John Benjamins. 176–186.
- Bernardini, Silvia. 2006. Corpora for translation education and translation practice: Achievements and challenges. *Third International Workshop on Language Resources for Translation Work, Research & Training*. 17–22.
- Bernardini, Silvia, Marco Baroni, and Evert Stefan. 2006. A Wacky introduction. *Wacky! Working papers on the web as corpus*, edited by Marco Baroni, and Silvia Bernardini. Bologna: GEDIT. 1–32.
- Bowker, Lynne. 1998. Using specialized monolingual native-language corpora as translation resource: a pilot study. *Meta: Translator's Journal* 43 (4). 631–651. <https://doi.org/10.7202/002134ar>.

- Bowker, Lynne. 2004. Corpora resources for translators: academic luxury or professional necessity. *Tradterm* 10. 213–247. <https://doi.org/10.11606/issn.2317-9511.tradterm.2004.47178>.
- Channell, Joanna. 2000. Corpus-based analysis of evaluative lexis. *Evaluation in Text*, edited by Susan Hunston, and Geoffrey Thompson. Oxford: Oxford University Press.
- Fantinuoli, Claudio. 2013. Projekte und Projektionen in der translatorischen Kompetenzentwicklung. *Einbindung von Korpora im Übersetzungsunterricht als Schlüssel zur Professionalisierung*, edited by Silvia Hansen-Schirra, and Don Kiraly. Frankfurt: Peter Lang. 173–188. <https://doi.org/10.3726/978-3-653-02047-2/12>.
- Fantinuoli, Claudio. 2016. Revisiting Corpus Creation and Analysis Tools for Translation Tasks. [www.periodicos.ufsc.br](http://www.periodicos.ufsc.br)
- Fantinuoli, Claudio. 2019. The Technological Turn in Interpreting: The Challenges that Lie Ahead. *Proceedings of the Conference Übersetzen und Dolmetschen 4.0. – Neue Wege in Digitalen Zeitalter*. 1–17.
- Francis, W. Nelson, and Kučera Henry. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston: Houghton Mifflin.
- Gatto, Maristella. 2014. *Web as Corpus. Theory and Practice*. London, New Delhi, New York, Sydney: Bloomsbury.
- Henzinger, Monika, and Steve Lawrence. 2004. Extracting knowledge from the World Wide Web. *Proceedings of the National Academy of Sciences* 101:5186–5191. <http://www.pnas.org/content/101/suppl.1/5186.full>. <https://doi.org/10.1073/pnas.0307528100>.
- Kilgarrieff, Adam. 2001. Web as Corpus. *Proceedings of the Corpus Linguistics 2001 Conference*, edited by Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, and Shereen Khoja. 342–344. UCREL, Lancaster. <http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/kilgarri.pdf>.
- Kilgarrieff, Adam, and Gregory Grefenstette. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics* 29 (3). 333–347. <https://doi.org/10.1162/089120103322711569>.
- Kübler, Natalie. 2011. Working with Different Corpora in Translation Teaching. *New Trends in Corpra and Language Learning*, edited by Ana Frankenberg-Garcia, Lynne Flowerdew, and Guy Aston. London: Continuum. 62–80.
- Kučera, Henry. 1991. The Odd Couple: The Linguist and the Software Engineer. The Struggle for High Quality Computerized Language Aids. *Directions in Corpus Linguistics – Proceedings of Nobel Symposium 82*, edited by Jan Svartvik. The Hague: Mouton de Gruyter. 401–419. <https://doi.org/10.1515/9783110867275.401>.
- Marcinkevičienė, Rūta. 2000. Tekstynų lingvistika (teorija ir praktika). *Darbai ir dienos* 24. 7–64.
- Marcinkevičienė, Rūta. 2002. Palyginamieji tekstynai – šaltinis tarptautinių žodžių vartosenai tirti. *Kalbotyra* 51(3). 81–93.
- McEnery, Tony, and Andrew Wilson. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Moon, Rosamund. 2010. What can a corpus tell us about lexis? *The Routledge Handbook of Corpus Linguistics*, edited by Anne O’Keefe, and Michael McCarthy. Routledge: New York. 197–211. <https://doi.org/10.4324/9780203856949.ch15>.
- Sinclair, John. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, John. 1996. The Search for Units of Meaning. *Textus* IX 1. 75–106.
- Stubbs, Michael. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Stubbs, Michael. 2009. The Search for Units of Meaning: Sinclair on Empirical Semantics. *Applied Linguistics* 30 (1). 115–137.

- Teich, Elke. 2003. *Cross-linguistic Variation in System and Text. A Methodology for the Investigation of Translations and Comparable Texts*. New York: de Gruyter. <https://doi.org/10.1515/9783110896541>.
- Teubert, Wolfgang. 1996. Comparable or Parallel Corpora? *International Journal of Lexicography* 9 (3). 238–264.
- Tognini-Bonelli, Elena. 2000. Corpus Currency. *Darbai ir dienos* 24. 205–243.
- Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Vaičėnionienė, Jurgita, Jolanta Kovalevskaitė, ir Teresė Ringailienė. 2017. Tekstynais paremti vertimų kalbos tyrimai ir šaltiniai. *Kalbų studijos* 30. 42–55.
- Varantola, Krista. 2003. Translators and Disposable Corpora. *Corpora in Translator Education*, edited by Federico Zanettin, Silvia Bernardini, and Dominic Stewart. Manchester: St Jerome. 55–70.
- Verplaetse, Heidi, and An Lambrechts. 2019. Surveying the use of CAT tools, terminology management systems and corpora among professional translators: general state of the art and adoption of corpus support by translator profile. *Parallèles* 31 (2).
- Zanettin, Federico. 2002. DIY corpora: the WWW and the translator. *Training the Language Services Provider for the New Millennium*, edited by Belinda Maia, Jonathan Haller, and Margherita Ulrych. Universidade do Porto. 239–248.
- Zanettin, Federico. 2012. *Translation-Driven Corpora*. Manchester: St. Jerome.
- Zipf, George K. 1935. *The Psycho-Biology of Language*. New York: Houghton Mifflin.