

Retrospektyvus Lietuvos sakytinės žiniasklaidos tekstynas: rengimo metodologija ir patirtis

Laima Nevinskaitė

Lietuvių kalbos institutas

laima@sociolingvistika.lt

Anotacija

Straipsnio tikslas – pristatyti akademinėi bendruomenei projekte „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“ parengtą apie 60 val. sakytinės žiniasklaidos kalbos tekstyną, aptarti metodologinius jo kūrimo principus ir problemas, pasidalinti praktine tekstų paieškos ir atrankos patirtimi. Straipsnyje pateikiama metodologinė tekstynų kūrimo principų apžvalga, daugiausia dėmesio skiriant reprezentatyvumo ir balanso kriterijams ir jų įgyvendinimo problematikai. Pristatoma teorinė tekstyno sandaros schema, besiremianti laiko (skirstymu į laikotarpius pagal sakytinės žiniasklaidos raidos etapus) ir žanro (išskirtos trys žanrinės grupės) kriterijais. Aptariami praktiniai ir metodologiniai šios schemos įgyvendinimo sunkumai, kuriuos lėmė ribotas medžiagos prieinamumas, medžiagos balansas tarp laikotarpių, žanrų palyginamumas tarp skirtingų laikotarpių, nevienoda žanrų įvairovė skirtingais laikotarpiais ir žanrų tęstinumas. Straipsnio pabaigoje pristatoma faktinė tekstyno sandara, apimtis ir techniniai jo požymiai.

Reikšminiai žodžiai: tekstynas, viešoji kalba, sakytinė žiniasklaida, sakytinės žiniasklaidos žanrai, tekstyno sudarymas.

Keywords: language corpus, spoken public language, spoken mass media, genres of spoken mass media, corpus sampling.

(Straipsnis parengtas vykdant projektą „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai, 2010–2013“ (vykdo Lietuvių kalbos institutas, vadovė Loreta Vaicekauskienė, remia Lietuvos mokslo taryba, sutarties Nr. VAT-14/2010).)

1. Įvadas

Tekstynai yra tapę viena iš svarbiausių kalbos analizės priemonių. Tačiau jų kūrimas dažnai reikalauja daug laiko ir finansinių resursų, ypač kai į kuriamą tekstyną norima įtraukti retrospektyvią medžiagą.

Lietuvių kalbos instituto įgyvendinamas projektas „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“¹ suteikė puikią progą parengti retrospektyvų Lietuvos sakytinės žiniasklaidos tekstyną. Vienas iš projekto tiriamųjų klausimų – ištirti kalbos idealo realizacijas viešojoje vartosenoje ir sakytinės žiniasklaidos kalbos kaitą, keliant hipotezę, kad lietuviškoje žiniasklaidoje vyksta panašūs kaip ir visoje Europoje stebimi viešosios kalbos kolokvializacijos arba konversacionalizacijos procesai (daugėja neformalios kalbos bruožų). Šiam klausimui atsakyti projekto metu buvo sukurtas 1960–2010 m. Lietuvos sakytinės žiniasklaidos (radijo ir televizijos) tekstynas (toliau tekste – Sakytinės žiniasklaidos tekstynas), šiuo metu apimantis 63 val. išrašytų, morfologiškai ir pagal pasirinktus kalbos bruožus anotuotų, su garsu/vaizdu susietų televizijos ir radijo įrašų.

Šiuo straipsniu norima pristatyti akademinėi bendruomenei minėtame projekte parengtą Sakytinės žiniasklaidos tekstyną, aptarti metodologinius jo kūrimo principus ir problemas, pasidalinti praktine tekstų atrankos ir rinkimo patirtimi. Tikimasi, kad tekstyno kūrimo principų ir patirties pristatymas suteiks idėjų kuriantiems panašaus pobūdžio tekstynus ateityje, leis atskleisti bendresnes tokio pobūdžio medžiagos rinkimo problemas, reflektuoti technologijų įtaką žiniasklaidos medžiagos saugojimui ir prieinamumui ir paskatins apmąstyti žiniasklaidos kalbos tekstų kaupimo perspektyvas ateityje. Straipsnis taip pat bus naudingas su tekstynu ateityje dirbsiantiems tyrėjams, nes leis geriau suprasti jo sudarymo principus ir panaudojimo galimybes bei apribojimus. Straipsnis skirtas aptarti tekstyno struktūros, tekstų atrankos klausimams, šį kartą paliekant nuošalyje išrašymo ir kodavimo problemas, kurios nebūtų naujos lyginant su kitų sakytinės kalbos tekstynų rengimo patirtimi (plg. Dabašinskienė ir Kamandulytė 2009).

Sakytinės žiniasklaidos tekstynas yra pirmasis tokio pobūdžio specializuotas tekstynas Lietuvoje. Bene geriausiai žinomame ir viešai prieinamame Vytauto Didžiojo universiteto (VDU) Kompiuterinės lingvistikos centro sukurtame Dabartinės lietuvių kalbos tekстыne² kaupiami tik rašytinės žiniasklaidos tekstai. Sakytinės žiniasklaidos kalbos išrašų taip pat yra VDU su partneriais parengtame Sakytinės lietuvių kalbos tekстыne³, tačiau jo tikslas buvo veikiau atspindėti sakytinės kalbos stilių ir žanrų įvairovę, tiesiogiai žiniasklaidai jis neskirtas (iš 230 šio tekstyno

¹ Lietuvos mokslo tarybos remiamas projektas „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“ (2010–2013 m.). Žiniasklaidos ir viešosios kalbos raidos tyrimas buvo viena iš trijų projekto dalių. Tekstyną rengiančiai darbo grupei vadovavo Laima Nevinskaitė, grupę sudarė Jurgita Girčienė, Giedrius Tamaševičius, konsultantai Žygintas Pečiulis ir Loreta Vaicekauskienė, dėl morfologinio anotavimo konsultavo Laura Kamandulytė-Merfeldienė, techniniais klausimais – Skaistė Aleksandravičiūtė.

² Vytauto Didžiojo universitetas, Kompiuterinės lingvistikos centras. Dabartinės lietuvių kalbos tekstynas [paieškos puslapis]. <http://tekstynas.vdu.lt/tekstynas/>. Žiūrėta 2013-07-23.

³ Vytauto Didžiojo universitetas. Sakytinės kalbos tekstynas [paieškos puslapis]. <http://donelaitis.vdu.lt/~andrius/SKT5/5SKT-paiesk.php>. Žiūrėta 2013-07-23.

reprezentuojamų sakytinės kalbos valandų, žiniasklaidos dalis sudaro apie 10 val. pokalbių laidų ir apie 16 val. žinių laidų; žr. Dabašinskienė ir Kamandulytė 2009). Čia pristatomas tekstynas išsiskiria ne tik gerokai didesne išrašytų laidų apimtimi, bet ir retrospektyvumu (sistemiškai reprezentuojama penkių dešimtmečių medžiaga) bei reprezentatyvia žiniasklaidos žanrų atranka (tekstynas reprezentuoja pokalbių, publicistines ir informacines laidas). Kiek žinoma, daugiau sistemingai kauptų reikšmingo dydžio tekstynų, kuriuose būtų reprezentuojama lietuviškoji sakytinė žiniasklaida, nėra.

Apie didesnės apimties specializuotus sakytinės žiniasklaidos tekstynus užsienio šalyse informacijos nepavyko rasti; tikėtina, kad tokia medžiaga yra įtraukiama į bendruosius tekstynus arba mažesne apimtimi kaupiama tyrėjų ar jų grupių savo tikslams. Minėto Dabartinės kalbos tekstyno puslapyje pateikiamame Užsienio tekstynų sąrašė minima nemažai tekstynų, įtraukiančių rašytinės žiniasklaidos medžiagą, tačiau sakytinė žiniasklaida nė prie vieno iš tekstynų nėra paminėta⁴.

Šiame straipsnyje iš pradžių pateikiama trumpa metodologinė tekstynų rengimo problematikos apžvalga, tada pristatoma teorinė Sakytinės žiniasklaidos tekstyno atrankos schema, aptariami metodologiniai ir praktiniai šios schemos realizavimo sunkumai ir galiausiai pristatoma faktinė parengto tekstyno struktūra bei jo techniniai požymiai. Rengiant straipsnį, naudotasi 2010 m. projekto ataskaitai rengtu „Sakytinės viešosios kalbos tekstų atrankos aprašu“ (Girčienė, Nevinskaitė ir Tamaševičius 2010).

2. Tekstynų sudarymas: metodologinė apžvalga

Tekstyną galima apibrėžti kaip baigtinį elektroninių tekstų rinkinį, sudarytą taip, kad kuo geriau atspindėtų kalbą ar jos atmainą (McEnery ir Wilson 1996: 177, cit. pagal Marcinkevičienė 2000). Tekstynas skiriamas nuo archyvų ar tekstinių duomenų bazių (tekstynas tekstai dažniausiai yra atrinkti taip, kad reprezentuotų tam tikrą kalbos atmainą ar žanrą, taip pat daugiau ar mažiau apdoroti, rengiant juos analizei, pgl. Baker, Hardy ir McEnery 2006) ir tiesiog tekstų (tekstynas yra kompiliacinės prigimties, tai yra tekstų rinkinys ir neanalizuojamas kaip atskiras tekstas su pradžia ir pabaiga (Marcinkevičienė 2000)).

⁴ Kompiuterinės lingvistikos centras. Kiti tekstynai. http://donelaitis.vdu.lt/main.php?id=4&nr=1_3. Žiūrėta 2013-07-23.

Literatūroje išskiriama daug tekstynų tipų pagal įvairius kriterijus (pgl. Marcinkevičienė 2000, Kennedy 1998, Hunston 2008; Baker, Hardy ir McEnergy 2006):

- bendrieji ir specializuoti tekstynai;
- dideli ir maži tekstynai;
- baigtiniai (statiniai) ir tęstiniai (dinaminiai) tekstynai;
- sinchroniniai ir diachroniniai tekstynai; Baker, Hardy ir McEnergy (2006) sinchroninius tekstynus tapatina su baigtiniais, o diachroninius su tęstiniais, nes pastarieji yra skirti kalbos raidai fiksuoti ir turi būti nuolat pildomi naujausia medžiaga (anglų k. dinaminiai tekstynai dar vadinami *monitor corpus*, nurodant jų paskirtį fiksuoti, stebėti kalbos kaitą);
- oportunistiniai (įtraukiantys bet kokius prieinamus tekstus) ir subalansuoti-reprezentatyvūs tekstynai;
- anotuoti ir koduoti bei neanotuoti tekstynai.

Dėl šių tipų privalumų ir trūkumų diskutuojama, o kai kurių iš jų pasirinkimas labiausiai priklauso nuo išorinių priežasčių, pvz., prieinamų resursų. Bene mažiausiai diskutuojama dėl subalansuotų-reprezentatyviųjų tekstynų pirmenybės oportunistinių tekstynų atžvilgiu, nes pripažįstama, jog tekstyno vertė labai priklauso nuo to, kiek apgalvota ir metodologiškai pagrįsta yra jo struktūra. Šis straipsnis ir yra skirtas pirmiausiai Sakytinės žiniasklaidos tekstyno sandarai, jo reprezentatyvumui aptarti, todėl reprezentatyvumo ir balanso kriterijai bus aptarti išsamiau, o kiti kriterijai tik trumpai paminėti toliau.

Tekstyno reprezentatyvumas apibrėžiamas kaip santykis tarp tekstyno ir kalbos tekstų visumos, kurią jis turėtų atspindėti (Hunston 2008: 160). Reprezentatyvus tekstynas atspindi tekstų visumos įvairovę (Biber 1993), remiantis tokiu tekstynu galima daryti apibendrintus teiginius apie kalbą arba tą jos dalį, kurią tekstynas reprezentuoja (Kennedy 1998). Taigi reprezentatyvumas yra būtina tekstyno vertingumo sąlyga – kuo labiau pagrįstus apibendrintus teiginius galima daryti remiantis tekstyno medžiaga, tuo vertingesnis yra tekstynas (Hunston 2008: 160). Reprezentatyvumo klausimas yra ypač sudėtingas kalbant apie bendruosius tekstynus, kuriems keliama užduotis atspindėti visą kalbą; specializuoto tekstyno atveju jis priklauso nuo tyrimo tikslo ir gali būti kiek paprastesnis.

Kitas kriterijus, į kurį turi būti atsižvelgiama sudarant tekstyną, yra balansas. Balansas apibūdina vidinę tekstyno sandarą, t. y. jo dalių (skirtingo tipo tekstų) proporcijas. Jei tekstyne yra

neproporcingai daugiau vienos rūšies tekstų negu kitos, tekstynas yra nesubalansuotas. Balanso reikalavimas (lygus pasiskirstymas tarp tekstyno dalių) gali kirstis su reprezentatyvumo reikalavimu (tekstyno dalių dydis priklausomai nuo jų reikšmės tyrinėjamoje kalbos visumoje). Subalansuoto tekstyno privalumas yra tai, kad kiekviena iš jo dalių yra pakankamai didelė, kad jas būtų galima patikimai lyginti tarpusavyje. Iš dalies balanso problemos sprendžiamos, normalizuojant skaičiuojamus kalbos požymių dažnius, nors vis tiek dalių dydis turi būti pakankamas, kad jas galėtumėm laikyti reprezentatyviomis ir iš jų daryti apibendrintas išvadas (Hunston 2008: 163–164).

Siekiant tekstyno reprezentatyvumo ir balanso, dažnas sprendimas yra stratifikuota atranka, kai išskiriami tam tikri objektyvūs tekstų požymiai ir juos derinant sudaromos tekstų grupės; tada tekstai atrenkami taip, kad pagal norimas proporcijas atspindėtų tas grupes (Biber 1993: 244). Tokiu atveju tekstyną galima traktuoti kaip mažesnių gana homogeniškų tekstynų rinkinį; dažniausiai taip sudaromi visi tekstynai, išskyrus labai specializuotus (Hunston 2008: 154). Tais atvejais, kai tekstynui nekeliamas tikslas apimti kokį nors itin specifinį kalbos atvejį (konkretaus žodžio ar gramatinės formos pavartojimą), tekstai dažnai yra atrenkami remiantis išoriniais kriterijais, t. y. naudojant situacinius, o ne kalbinius tekstų požymius, pvz., žanrą ar registrą (Biber 1993: 244–245). Šių atrankos principų buvo laikomasi ir pristatomo Sakytinės žiniasklaidos tekstyno atrankoje, kur buvo derinami laikotarpio ir žanro požymiai (žr. toliau).

Kuriant specializuotus tekstynus, aktualius tekstų tipus ar grupes lemia tekstyno kūrimo tikslas, tačiau literatūroje išskiriami ir bendrieji tekstų atrankos kriterijai, kurie gali būti aktualūs ir specializuotų tekstynų kūrėjams. Pavyzdžiui, literatūroje minimi tokie rašytinių tekstų atrankos kriterijai: elitiškumas (tekstų „įtakingumas“); atsitiktinė atranka; paplitimas (tekstų populiarumas, pvz., pagal šį kriterijų teikiama pirmenybė žurnalistiniams tekstams ir populiariausioms knygoms); subjektyviai įvertinamas „tipiškumas“; tekstų prieinamumas; demografiniai skaitomumo duomenys; tekstų pasirinkimas pagal kalbinius požymius. Kaip teigia autorius, pragmatinis sprendimas paprastai apima šių įvairių kriterijų derinį (Summers 1991, cit. pagal Kennedy 1998: 63–64).

Kai tekstynas kuriamas konkrečiam tyrimui, kaip buvo Sakytinės žiniasklaidos tekstyno atveju, tekstų atrankos kriterijus pirmiausia lemia tyrimo tikslai, jo parametrai. Jie gali būti formuluojami remiantis kalbos variantiškumo arba kitomis teorijomis. Vis tik, kaip teigia Hunston (2008: 155) kintamųjų sąrašas dažniau yra panašus į sveiku protu paremtą klasifikaciją negu vienos kurios

teorijos rezultata, išskiriant tokius intuityviai svarbus situacinius požymius, kaip monologas ar dialogas, tiesioginis pokalbis ar pokalbis telefonu ir pan.

Tekstyno reprezentatyvumui svarbus klausimas yra ir tekstyno (t. y. atrinktų tekstų imties) dydis. Nors yra pateikiama metodų tekstynų imties dydžiui apskaičiuoti (žr. Biber 1993), jie yra skirti ir yra aktualūs rašytinės kalbos tekstynams, kur gali būti įmanoma pasiekti imties reprezentatyvumą dydžio atžvilgiu. Sakytinės kalbos tekstynų reprezentatyvumas negali būti matuojamas pagal tuos pačius kriterijus. Šie tekstynai dėl jų parengimui reikalingų didelių darbo ir laiko sąnaudų dažniausiai yra palyginti maži, ypač jei jų išrašai yra labai tikslūs ir anotuoti. Be to, dėl būtinybės planuoti tekstyno parengimui reikalingas sąnaudas, dažnai tekstyno dydis yra apibrėžiamas iš anksto ir priklauso nuo prieinamų resursų (Hunston 2008: 165–166).

Dar vienas iš Hunston (2008: 162) nurodytų būtų spręsti bendrųjų tekstynų reprezentatyvumo problemas, svarbus kalbant ir apie nedidelės apimties specializuotus tekstynus, tai – „pripažinti reprezentatyvumo problemas ir skaidriai aprašyti, kaip buvo sukurtas tekstynas ir kas įtraukta į jį“. Tai leidžia tekstyną naudojančiams tyrėjams patiems įvertinti jo reprezentatyvumą ir atitinkamai pasirinkti medžiagos panaudojimo būdą ar interpretuoti išvadas. Tas pats principas taikomas ir kalbant apie tekstyno balanso problemas – tekstyno sandara ir šiuo atžvilgiu turi būti skaidriai aprašyta. Ne visiškai subalansuotas tekstynas „nekelia sunkumų, jei tai yra žinoma ir galima atitinkamai pakoreguoti hipotezes“ (Hunston 2008: 163). Būtent tokio skaidrumo ir siekiama šiame straipsnyje, aprašant tekstyno kūrimo problemas ir apribojimus.

Tačiau autorė taip pat pažymi, kad išsamus tekstyno aprašymas galimas tik iš dalies ir argumentai dėl kurio nors tekstyno reprezentatyvumo ir subalansuotumo yra neišvengiamai tautologiški, nes kategorijos, pagal kurias aprašoma tekstyno sandara, yra nulemtos jo dizaino. Kitaip sakant, tos kategorijos, kuriomis buvo remiamasi sudarant tekstyną, gali būti reprezentatyvios ir subalansuotos, tačiau gali būti tokių kategorijų, kurios yra mažiau pastebimos (kurios nebuvo iškeltos kaip tekstyno sudarymo kriterijus) ir jų atžvilgiu tekstynas gali būti nereprezentatyvus (Hunston 2008: 164).

Neskaitant tekstų atrankos sunkumų, susijusių su tekstynų reprezentatyvumu ir balansu, tekstynų kūrėjams iškyla ir praktinių sunkumų. Hunston žodžiais (2008: 156), „visi tekstynai yra kompromisas tarp to, kas siekiama, ką planavo tekstyno kūrėjas, ir kas yra įmanoma“. Kuriant tekstynus, susiduriama su daugeliu praktinių apribojimų, tarp kurių minimi apribojimai dėl

programinės įrangos, autorių teisių, etinių klausimų ir tekstų prieinamumo. Kuriant čia pristatomą Sakytinės žiniasklaidos tekstyną, didžiausias apribojimas buvo tekstų prieinamumas. Tačiau taip pat kilo sunkumų, kuriuos lėmė metodologiniai pristatomos schemos aspektai. Visi jie bus aptariami toliau, pristatčius teorinę tekstyno atrankos schemą.

3. Teorinė tekstyno atrankos schema

Sakytinės žiniasklaidos tekstynas yra specializuotas tekstynas. Jis apima gana siaurai apibrėžtą tekstų tipą – ne menines televizijos ir radijo laidas. Tai yra retrospektyvus, arba diachroninis tekstynas. Jis sukauptas siekiant atsakyti į klausimus, keltus straipsnio įvade minėtame projekte, t. y. ištirti kalbos idealo realizacijas viešojoje vartosenoje ir sakytinės viešosios kalbos kaitą, ypač tikrinant teorinę prielaidą apie kalbos neformalėjimą. Viešosios kalbos pokyčius buvo siekiama susieti su platesniu viešosios erdvės pokyčių kontekstu. Atitinkamai buvo išskirti kriterijai, pagal kuriuos modeliuota tekstyno sandara.

3.1. Sakytinės žiniasklaidos tekstyno sudarymo kriterijai

Kuriant tekstyno atrankos schemą, buvo remiamasi teorinėmis prielaidomis apie sakytinės žiniasklaidos raidą ir jos įtaką kalbai sakytinėje žiniasklaidoje, iš karto bandant atsižvelgti į realią Lietuvos žiniasklaidos situaciją bei galimas laidų prieinamumo problemas. Taigi sudarant schemą, remtasi žiniasklaidos raidą ir pokyčius įvairiais aspektais (tarp jų ir kalbiniu) nagrinėjančiais tyrimais (žr. Fairclough 1995; Hedin 2005; Pečiulis 2007; Tamaševičius 2009; 2010; Girčienė 2009; 2010 ir kt.) ir retrospektyvine programų tinklelių bei viešai prieinamų archyvų peržvalga.

Siekiant tekstyno reprezentatyvumo ir balanso, jo atrankos schema rėmėsi dviem kriterijais: laiko (tekstų atranka pagal sąlyginius sakytinės žiniasklaidos kaitos etapus Lietuvoje) ir žanrų (atranka pagal laidų pobūdį). Šiėdu pagrindiniai kriterijai išsamiau pristatomi toliau.

3.2.1. Laidų atranka pagal sakytinės žiniasklaidos raidos etapus

Siekiant analizuoti kalbos kaitą pirmiausia buvo būtina išskirti etapus, arba laikotarpius, kuriais remiantis būtų galima lyginti kalbą skirtingais laikotarpiais. Kalbos pokyčiai neišvengiamai susiję su žiniasklaidos ir apskritai viešosios erdvės pokyčiais, todėl buvo išskirti sakytinės žiniasklaidos raidos etapai, kurie naudojami ir laidų atrankai, ir vėliau analizuojant kalbos kaitą.

Sakytinės žiniasklaidos raidos etapai išskirti atlikus retrospekcinę žiniasklaidos raidos apžvalgą nuosavybės formų, komunikacijos pobūdžio, žanrinės įvairovės, kalbėtojų, adresato, auditorijos įtraukimo ir kt. atžvilgiais. Išskirti trys pagrindiniai sakytinės žiniasklaidos (televizijos ir radijo) raidos Lietuvoje etapai:

1. Sovietinis laikotarpis: 1960–1987 m. Laikotarpio pradžios data buvo parinkta palyginti sąlygiškai, siejant su didesniu televizijos plitimu Lietuvoje (televizijos transliacijos pradėtos 1957 m.) ir su projekto tikslais – tirti paskutinių penkių dešimtmečių kalbos ir standartizavimo ideologijų raidą.

2. Pereinamasis laikotarpis: 1988–1992 m. 1988 metai pasirinkti kaip pereinamojo sakytinės žiniasklaidos kaitos laikotarpio pradžia, nes tais metais vyko ryškesni pokyčiai viešojoje erdvėje (plg. Nevinskaitė 2006) ir pasirodė pirmosios „naujosios kartos“ televizijos laidos, tarp kurių labiausiai minėtinas „Veidrodis“ (Pečiulis 2000)⁵.

3. Dabartinis laikotarpis: 1993–2011 m. Jo pradžia siejama su pirmosios komercinės televizijos TELE-3 įsteigimu. Nors kalbant apie tekstyną nurodoma, kad jis apima penkis dešimtmečius (1960–2010), iš tiesų tekstyno rengėjai pasinaudojo proga dabartinį laikotarpį papildyti pačiomis naujausiomis laidomis, taigi dabartinis laikotarpis tekстыne apima ir 2011 metus.

Šiems laikotarpiams būdingi skirtingi požymiai, jų derinys galiausiai lemia ir žiniasklaidos kalbos pobūdį. Laikotarpių ypatumai apibendrintai parodyti 1 lentelėje.

Raidos bruožai / Etapas	Sovietinis laikotarpis (1960–1987)	Pereinamasis laikotarpis (1988–1992)	Dabartinis laikotarpis (1993–2011)
Nuosavybės forma	Valstybinė	Valstybinė, komercinės užuomazgos	Komercinė ir visuomeninė
Vaidmuo	Šviečiamoji funkcija, „visažinio mokytojo“ vaidmuo	Šviečiamoji, informavimo funkcija, kolegiškumas	Pramoginė funkcija, (imituojama) interakcinė draugiška komunikacija

⁵ „Veidrodis“ tapo „populiariausia ir aštriausia Atgimimo metų Lietuvos televizijos laida [...]. Joje pirmą kartą pasirodė daugelis politikų, Sąjūdžio įkūrėjų. „Veidrodis“ pirmasis prašneko apie Rainių tragediją, kruvinočius Tbilisio įvykius. Šioje laidoje pirmą kartą per TV kalbėjo kunigas, pirmą kartą buvo viešai parodytas tremtinių susitikimas. Buvo išmėgintos ir technologinės naujovės. „Veidrodis“ po ilgų cenzūravimo metų sugrąžino tiesiogines transliacijas, pradėti rengti TV tiltai, apklausos telefonu. [...] Laidos sėkmę lėmė atviras kalbėjimo būdas, kova prieš tuometinę nomenklatūrą, pasišaipymas iš valdžios“ (Pečiulis 2000: 233–234).

Žanrai	Vyrauja informacinės laidos ir publicistiniai žurnalai	Informacinės laidos ir publicistinius žurnalus papildo daugiau ar mažiau spontaniškų pokalbių laidos	Vienodai svarbios informacinės, publicistinės ir pokalbių laidos, didelė žanrų ir požanrių įvairovė
Komunikacijos kryptis	Vienakryptė, dominuoja profesionalūs kalbėtojai	Randasi grįžtamojo ryšio užuomazgų, neprofesionalių kalbėtojų	Neprofesionalūs kalbėtojai konkuruoja su profesionalais
Dalyvių skaičius	Vyrauja monologas	Randasi daugiau dialoginio kalbėjimo	Vyrauja dialogas
Kalbos pobūdis	Vyrauja parengta nespontaniška kalba	Randasi neparengto spontaniško kalbėjimo	Vyrauja neparengtas spontaniškas kalbėjimas

1 lentelė. Kalbos analizei reikšmingi radijo ir televizijos raidos etapai (Girčienė, Nevinskaitė ir Tamaševičius, 2010, remiantis Pečiulis 2007)

Toliau straipsnyje dar bus aptarta, kaip pavyko įgyvendinti laidų atranką pagal šiuos laikotarpius ir kaip pasiteisino toks laikotarpių išskyrimas.

3.2.2. Žanrinis laidų atrankos kriterijus

Tekstų pobūdžio atžvilgiu sakytinės žiniasklaidos tekstų atranka siejama su žanrais – laidų tipais, apibūdinančiais to tipo laidoms būdingus bendrus požymius (Bignell 2004). Žanras yra gana stabili tekstinė struktūra, palaikanti komunikacijos tęstinumą ir darną (Lüders et al. 2010). Taigi atliekant tekstų atranką, naudinga ir net neišvengiamai tenka remtis laidų žanrais, nes jie leidžia numatyti kalbos tyrimo aspektu svarbius laidų požymius (pvz., spontaniškumą, dialogiškumą). Kitaip sakant, žanrai atrankoje tarnavo kaip tam tikras išorinis kalbos atžvilgiu pragmatinis kriterijus, leidžiantis paprasčiau atrinkti laidas, išsamiai neanalizuojant kiekvienos laidos atskirai. Tačiau ir tekstų skirstymas į žanrus nėra visiškai objektyvi procedūra, todėl ir šiuo atveju tekstynų kūrėjams rekomenduojama išsamiai aprašyti žanrų kategorijas, kad tekstyno naudotojai aiškiai suprastų, ką apima vienas ar kitas žanras (Baker, Hardy ir McEnery 2006).

Daugiausia dėmesio skirta televizijos laidų žanrams, nes dėl archyvinių įrašų trūkumo į tekstyno imtį buvo įtraukiama gana nedaug radijo tekstų. Išskiriant ir apibūdinant žanrus remtasi prancūzų autoriaus Laurento Fonneto šiuolaikinių laidų klasifikacija, perteikta Žyginto Pečiulio (2007), pakoreguojant ir praplečiant ją pagal dabartinę Lietuvos televizijų programų situaciją. Fonnetas laidas skirsto į daugkartinio (*stock*) ir vienkartinio (*flux*) demonstravimo – šis skirtumas bus aktualus svarstant laidų įrašų prieinamumą. Daugkartinio rodymo laidos, kaip sako pats jų

pavadinimas, yra didesnės išliekamosios vertės ir dažniau rodomos pakartotinai; vienkartinio rodymo laidos paprastai kuriamos konkretaus kanalo reikmėms; jos skirtos kasdieniniams televizijos poreikiams ir rečiau tiražuojamos (Pečiulis 2007). Konkretūs daugkartinio ir vienkartinio rodymo žanrai pateikiami 2 lentelėje.

Daugkartinio rodymo laidos (audiovizualiniai kūriniai)	Vienkartinio rodymo (kasdieninės) laidos
Kinematografo kūriniai (kino filmai) Televizijos filmai (serialai, muilo operos, situacijų komedijos) Dokumentika (dokumentiniai filmai ir išskirtinė tiriamoji žurnalistika) Animacija	Pramoga (varjetė, teatralizuoti koncertai, infošou) Žaidimai (proto, jėgos, mišrūs) Žurnalai (laidos, susidedančios iš atskirų dalių ir siejamos vedėjo) Pokalbių laidos (debatai, pokalbių šou) Realybės šou (tikrovę imituojančios laidos ir kt.) Patarimai (vartojimas, buitės, pomėgiai) Sportas (naujienos ir transliacijos) Informacija (žinios ir analitinės laidos) Jaunimo laidos (skiriasi tematika ir teisiniais apribojimais)

2 lentelė. Fonet televizijos laidų klasifikacija (pagal Pečiulis 2007: 140–141)

Ribota tekstyno apimtis (apie 60 val.) riboja reprezentuojamų laidų pobūdį, todėl kaip tinkamiausi atsakyti į tiriamąjį klausimą atrinkti tik suaugusiam žiūrovui skirti bendrosios tematikos dokumentiniai (atspindintys realius faktus ir įvykius) žanrai, kurių pagrindinę dalį sudaro viešajam **kalbėjimui** (ne vaidybai ar atlikimui) priskirtinas diskursas. Taigi atsisakyta imitacinio pobūdžio diskursą reprezentuojančių vaidybinių žanrų (vaidinimų, filmų, serialų ir kt., humoro laidų). Į imtį netraukiami realybės šou, reprezentuojantys labiau privačiam, o ne viešajam diskursui priskirtiną kalbėjimą, arba galintį turėti vaidybos elementų. Dėl stipraus vaidybinio pobūdžio į imtį netrauktas ir infošou žanras – programos, kuriose „nerimtai“, žaismingai ir su vaidybos elementais interpretuojamos aktualijos. Teminiu atžvilgiu į imtį įtrauktos bendrojo pobūdžio, t. y. visuomeninės politinės ir bendrajai auditorijai skirtos pramoginės laidos, atsisakyta specializuotų patarimų žanrui priskirtinų laidų (pvz., žvejams, apie automobilius ir kt.), taip pat specializuotai auditorijai (pvz., vaikams) skirtų laidų.

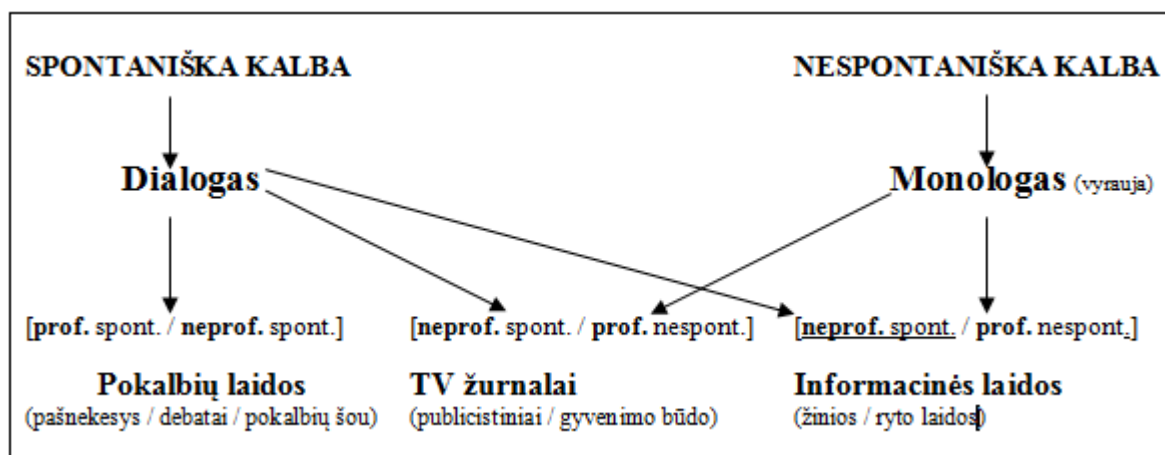
Siekiant tiksliau apibrėžti laidų atranką, pirmiausia buvo sudaryta projekto tyrimui svarbių sakytinės žiniasklaidos žanrų komunikacinių ir kalbinių bruožų schema (žr. 1 schemą). Schema paremta pagrindine kalbos ypatybes lemiančia dichotomija: gyvai produkuojama iš anksto neparengta spontaniška kalba vs iš anksto parengta, (dažniausiai) skaitoma, todėl nespontaniška,

kalba. Spontaniškas kalbėjimas būdingiausias dialogo ar polilogo dalyviams, nespontaniškas kalbėjimas paprastai esti monologinis. Šių skirtingų kalbos produkavimo būdų (samplaikų) realizacija sietina su trimis pagrindiniais sakytinės žiniasklaidos žanrais (žr. Hedin 2005; Pečiulis 2007) – pokalbių laidomis, žurnalais ir informacinėmis laidomis:

1. Pokalbių laidoms būdingas daugiau ar mažiau spontaniškas laidų vedėjų (profesionalių kalbėtojų) ir laidos svečių (paprastai neprofesionalių kalbėtojų) dialogas ar polilogas. Šioje grupėje išskirti keli smulkesni laidų tipai: pašnekesių, arba interviu, laidos, pokalbių šou ir debatai.

2. Publicistiniai ir gyvenimo būdo žurnalai – tai laidos, susidedančios iš kelių siužetų, siejamų ir pristatomų to paties vedėjo, pvz., „Prieš srovę“, „Nuo iki“ ir kt. Jų vedėjai, profesionalūs kalbėtojai, paprastai kalba nespontaniškai (dažnai skaito užkadrinį tekstą), o kalbinami neprofesionalūs kalbėtojai (laidų svečiai) kalba spontaniškai. Šioms laidoms būdingas nespontaniško monologinio ir spontaniško dialoginio kalbėjimo derinys.

3. Informacinių laidų vedėjai ir reportažų autoriai profesionalai paprastai kalba nespontaniškai (dažnai skaito tekstą), o reportažų rengėjų kalbinami neprofesionalai kalba spontaniškai (kadangi informacinėse laidose vyrauja vedėjų ir reportažų autorių kalba, tokiose laidose vyrauja nespontaniškas monologinis kalbėjimas).



1 schema. Tekstyno laidų žanrų komunikaciniai ir kalbiniai bruožai (Girčienė, Nevinskaitė ir Tamaševičius 2010) *Sutrumpinimai*: prof. – profesionalūs kalbėtojai, neprof. – neprofesionalūs kalbėtojai, spont. – spontaniškas kalbėjimas, nespont. – nespontaniškas kalbėjimas.

Visi aptartieji žanrai ir požanriai būdingi televizijos laidoms. Radijo laidų klasifikacija galėtų būti panaši į televizijos laidų, nors ir esama tam tikrų skirtumų. Dėl radijo laidų archyvinių įrašų

trūkumo į imtį buvo traukiama palyginti nedaug radijo laidų (žr. toliau), todėl radijo laidų klasifikacijos klausimai detaliau nenagrinėjami.

Žiniasklaidos laidos laikotarpiai ir žanrai buvo pagrindiniai kriterijai, atrenkant laidas į tekstyną. Tais atvejais, kai buvo galima rinktis iš kelių to paties žanro ir to paties laikotarpio laidų, buvo siekiama įtraukti žinomesnes, populiariesnes laidas, vedėjus, laidose kalbintas asmenybes, nors ir įvertinant tai subjektyviai, remiantis tekstyno kūrėjų patirtimi. Pavyzdžiui, į tekstyną pateko laidos su tokiais žinomomis televizijos ir radijo asmenybėmis kaip Algimantas Čekuolis, Alfredas Bumblauskas, Nomeda Marčėnaitė, Rūta Grinevičiūtė, Giedrius Masalskis ir kt.

4. Tekstyno sandaros apribojimai ir panaudojimo galimybės

Bandant įgyvendinti pristatytą teorinę atrankos schemą, susidurta su sunkumais, kuriuos lėmė ne tik praktiniai tekstų prieinamumo, bet ir metodologiniai aptartosios schemos aspektai. Jie susiję ir su sakytinės žiniasklaidos kaitos etapais, ir su išskirtais žiniasklaidos žanrais.

4.1. Sakytinės žiniasklaidos etapų reprezentavimas

Siekiant tinkamai reprezentuoti išskirtus sakytinės žiniasklaidos raidos etapus, didžiausia iškilusi problema buvo įrašų prieinamumas, kurį lėmė technologijų ypatumai ir archyvinių šaltinių organizaciniai veiksniai, aprašomi toliau. Skyriaus pabaigoje aptariamas laikotarpių ribų išskyrimo klausimas ir apibendrinama, kaip pavyko subalansuoti atspindėti išskirtus laikotarpius.

4.1.1. Technologijos ir archyvinės medžiagos prieinamumas

Atrenkant medžiagą pagal išskirtus sakytinės žiniasklaidos raidos etapus, didžiausi sunkumai iškilo dėl medžiagos prieinamumo. Skirtingą laidų iš skirtingų laikotarpių prieinamumą lėmė skirtingais laikotarpiais naudotų technologijų įtaka ir sąsaja su žanrinėmis laidų grupėmis: minėtomis daugkartinio rodymo ir vienkartinio rodymo laidomis.

Kaip ir buvo galima tikėtis, sunkiausiai prieinamos senesnių – sovietinio ir pereinamojo – laikotarpių laidos. Iš sovietinio laikotarpio rasta daugiau išlikusių tų laidų ar jų fragmentų, kurios buvo įrašomos į kino juostą; o į kino juostą būdavo įrašoma daugkartinio rodymo medžiaga – dokumentiniai filmai, apybraižos, taip pat meninis turinys (pvz., koncertai), kuris į tekstyną nebuvo traukiamas. Vėliau, pradėjus naudoti vaizdo įrašymo į magnetinės juostos kasetes technologiją, laidos buvo įrašomos „ant viršaus“, naujuoju įrašu ištrinant senąjį. Tos laidos ar laidų fragmentai,

kurie buvo transliuojami tiesiogiai, pvz., diktorių skaitomi žinių laidų fragmentai, dažniausiai nebūdavo įrašomi ir jų neišliko. Tai yra viena iš priežasčių, kodėl tekстыne iš sovietinio laikotarpio daugiausia įtraukta daugkartinio naudojimo dokumentinių apybraižų – nes jų yra daugiausiai išlikusių iš į imtį norimų įtraukti žanrų. Visuose tekstyno reprezentuojamuose laikotarpiuose pavyko įtraukti tik labai nedaug informacinių laidų, nes jų, kaip labiausiai kasdieninio naudojimo žanro, nebuvo įrašoma. Rečiau saugoti ir reportažų įrašai, jų skaičius tekстыne taip pat nedidelis.

Iš sovietinio laikotarpio yra išlikusių gana daug radijo įrašų, ir jais buvo kompensuotas sovietinės televizijos įrašų trūkumas; todėl tekстыne daugiausiai radijo įrašų yra iš sovietinio laikotarpio. Toks radijo prieinamumas gali būti paaiškinamas technologiškai, nes garso įrašymas į magnetines juostas paprastesnis ir pigesnis, negu vaizdo. Tik garso formatu išliko ir kelios televizijos laidos. Kita vertus, radijas sovietiniu metu atliko svarbesnę vaidmenį negu dabar, nes tuo metu transliavo tik vienas televizijos kanalas lietuvių kalba ir pora kanalų rusų kalba, bent jau pirmaisiais televizijos dešimtmečiais buvo sąlyginai sunkiau įsigyti televizorių ir matyti televizijos programas. Todėl didesnis radijo įrašų prieinamumas atspindi jo svarbą tuo metu ir taip galima pateisinti didesnę jam skiriamos apimties dalį.

4.1.2. Archyvinės medžiagos šaltiniai ir jų prieinamumas

Medžiagos tekstynui paieška archyvinuose šaltiniuose pareikalavo nemažai pastangų. Archyvinė radijo ir televizijos medžiaga yra išskaidyta per skirtingas saugojimo vietas (institucijas), labai trūksta susistemintos informacijos apie archyvinės medžiagos pobūdį ir apimtį, kai kuriais atvejais ši informacija apskritai nėra skelbiama ir pateikiama tik užklausus. Taigi Lietuvoje labai trūksta centralizuotų audiovizualinės archyvinės medžiagos tvarkymo ir prieinamumo kūrimo iniciatyvų, kurios turbūt ir turėtų kilti iš pačių tokios medžiagos prieinamumu suinteresuotų tyrėjų. Archyvinė medžiaga jau tampa prieinamesnė – pvz., į portalą e-kinas.lt jau yra sukelta ir viešai prieinama nemažai audiovizualinės medžiagos (kino kronikų ir kt., kurios taip pat gali būti aktualios kalbos analizei), tačiau jis orientuojamas į kino produkciją ir neapima radijo ir televizijos įrašų.

Čia galima nurodyti pavyzdį iš Danijos, kur vykdomas didelis tarpdalykinis projektas LARM, skirtas sukurti infrastruktūrą (garso įrašų duomenų bazę su papildomais įrankiais ir medžiagos aprašais), leidžiančią įvairių sričių (kalbos, žiniasklaidos, kultūros istorijos ir t. t.) tyrėjams patogiai skaitmenine forma naudotis visu Danijos radijo palikimu. Planuojama, kad joje bus prieinama

daugiau negu milijonas valandų radijo įrašų⁶. Nors dėl susistemintos informacijos trūkumo sunku įvertinti, kokia yra Lietuvos archyvuose saugomos audiovizualinės medžiagos apimtis, tekstyno sudarytojų įspūdžiu, ji yra labai maža, lyginant su paminėtais skaičiais.

Vis dėlto, nepaisant riboto įrašų prieinamumo, mūsų palyginti nedidelės apimties tekстыne pavyko gana reprezentatyviai atspindėti visus išskirtus laikotarpius. Skirtingų laikotarpių įrašai buvo renkami iš skirtingų šaltinių, pagrindiniai buvo:

LRT Aukso fondas, kuris šiuo metu nebeveikia, jo medžiaga yra integruojama į LRT Fonoteką. Šis rinkinys buvo vienas iš dviejų svarbiausių sovietinio ir pereinamojo laikotarpio laidų įrašų šaltinių; jame saugoma ir nemažai dabartinio laikotarpio Lietuvos nacionalinės televizijos laidų. LRT Aukso fondas internete leido tik žiūrėti laidas einamuoju laiku, tačiau maloniai bendradarbiaujant su LRT, buvo gauti reikiami fondui priklausančių laidų failai. Nors LRT Aukso fondas atrodė akivaizdžiausias archyvinės televizijos medžiagos šaltinis, jis nebuvo pakankamas, nes iš sovietiniam laikotarpiui reprezentuoti reikalingos medžiagos jame buvo rasta nemažai dokumentinių filmų ir apybraižų, bet praktiškai nebuvo pokalbių laidų (kaip pokalbių laida įtraukta tik 1987 m. „Sporto studija“), labai mažai – informacinių laidų. Be to, jame seniausias į tekstyną įtraukti tinkamas įrašas buvo tik iš 1967 m. Šiuos trūkumus padėjo kompensuoti kitas toliau aprašomas šaltinis.

Lietuvos valstybės centrinio archyvo Garso ir videodokumentų skyriuje sukaupta nemažai sovietinio laikotarpio radijo laidų ar jų fragmentų įrašų (archyvui buvo perduotas saugoti radijo archyvas kartu su leidimu jį naudoti), šiuo metu iš dalies suskaitmenintų. Būtent iš šio archyvo įrašų buvo labiausiai reprezentuotas septintasis dešimtmetis bei pildomos aštuntojo dešimtmečio spragos (naujesnių tekstynui tinkamų įrašų nebuvo). Be radijo laidų, Lietuvos valstybės centrinio archyvo rinkiniai tekstyno rengėjams pateikė kelias malonias staigmenas: archyve rasti du vienos iš nedaugelio sovietiniais metais egzistavusių televizijos pokalbių laidų „Televizijos forumas“ garso įrašai (1977 m.) ir dar kelių televizijos laidų garso įrašai. Taigi radijo ir televizijos laidų garso įrašais buvo kompensuotas sovietinio laikotarpio pokalbių ir informacinių televizijos laidų įrašų trūkumas. Be to, archyvas turi nedidelį mūsų išskirto pereinamojo ir ankstyvojo komercinio laikotarpio televizijos laidų rinkinį, sukaupimą daugiausia iš archyvo darbuotojų ar kitų asmenų padarytų įrašų, kurie taip pat buvo labai reikalingi tekstynui.

⁶ LARM Audio Research Archive. <http://www.larm-archive.org/>

Ieškant dabartinio laikotarpio komercinių televizijų laidų, buvo kreiptasi į pačias **komercines televizijas: LNK, TV3, Baltijos TV**. Čia taip pat susidurta su įrašų prieinamumo problema, nes iki skaitmeninių technologijų paplitimo ne visos laidos būdavo įrašomos, ne visi įrašai išlikę, ypač tiesiogiai transliuotų laidų. Dėl šios priežasties į tekstyną nepavyko įtraukti visų numatytų dabartinio laikotarpio laidų. Vis dėlto geranoriškas televizijų bendradarbiavimas ir suteiktas leidimas naudoti patiems padarytus įrašus leido dabartinį tekstyno laikotarpį reprezentuoti palyginti išsamiai.

Ieškant dabartinio laikotarpio radijo įrašų, taip pat buvo kreiptasi į **komercines radijo stotis M-1 ir Radiocentrą**. Jos teigė nekaupiančios įrašų archyvų. Nors daryti transliacijų įrašus įpareigoja įstatymai, pasibaigus trumpam įstatymų numatytam laikui, įrašai nėra kaupiami. Komercinio radijo archyvinių įrašų pavyko gauti tik iš vieno laidos vedėjo asmeninio archyvo bei su stočių leidimu patiems pasidaryti transliacijų įrašų.

Sakytinės žiniasklaidos tekstynui atrinkti įrašai naudojami su visų radijo ir televizijos stočių, per kurias jie buvo transliuojami, sutikimu. Pasinaudodami proga visų projekto vykdytojų vardu dėkojame bendradarbiavusioms televizijos ir radijo stotims: LTV, LNK, TV3, Baltijos TV, M-1, Radiocentrai, asmeninius įrašus suteikusiam Giedriui Masalskiui, paslaugoms Lietuvos valstybinio centrinio archyvo darbuotojoms.

4.1.3. Laikotarpių ribos ir medžiagos balansas

Kaip minėta, išskirtieji žiniasklaidos raidos laikotarpiai atriboti pagal formalius požymius, leidžiančius kelti prielaidas apie tais laikotarpiais dominavusius kalbos bruožus. Iš darbo rengiant tekstyno išrašus ir pradinės analizės matyti, kad kaita vyksta palaipsniui ir laikotarpių ribos yra iš dalies sąlyginės. Todėl ypač pasiteisino pereinamojo laikotarpio išskyrimas, nes paaiškėjo, kad ne tik pagal formalius žiniasklaidos raidos bruožus, bet ir pagal kalbos kaitos požymius jis iš tiesų yra pereinamasis tarp sovietinio ir dabartinio laikotarpio, jame galima matyti ir ankstesniojo, ir tolesniojo laikotarpių bruožų (pvz., žr. Aleksandravičiūtė ir Vaicekauskienė 2012; Girčienė ir Tamaševičius 2012; Čičirkaitė 2012).

Galima teigti, kad tekстыne pavyko gana išsamiai reprezentuoti visus tris išskirtus sakytinės žiniasklaidos raidos etapus. Remiantis pirmųjų tyrimų iš tekstyno medžiagos patirtimi matyti, kad

medžiagos užtenka atlikti palyginimą tarp skirtingų laikotarpių, o kartais net tarp skirtingų dešimtmečių. Tokį palyginamumą užtikrino ir vidinis laikotarpių balansas. Nors sudarant tekstyną buvo išskirti trys stambūs etapai ir pirmiausiai siekta kiekvieno iš jų reprezentatyvaus atspindėjimo tekстыne, buvo siekiama ir kiek įmanoma tolygesnio pasiskirstymo pagal metus. Galima sakyti, kad tokį balansą tekstyno rengėjams pavyko pasiekti – dėl archyvinės medžiagos trūkumo tekстыne yra tik keleri metai, kurių nereprezentuoja nė vienas įrašas.

Vis dėlto, pritrūkus medžiagos ar dėl kitų priežasčių norint sumažinti laikotarpių skaičių ir išskirti du pagrindinius etapus, skiriamąją riba laikytini 1988 m. – laikoma, kad pereinamasis laikotarpis struktūriškai panašesnis į dabartinį etapą ir jungtinas prie jo.

4.2. Žanrinio skirstymo problemos

Nors žanrinis skirstymas turėtų leisti lyginti skirtingus tekstus iš atskirų laikotarpių, tai nėra vienareikšmiška ir paprasta. Sudarant tekstyną, susidurta su žanrų palyginamumo, žanrų paplitimo skirtingais laikotarpiais ir žanrų tęstinumo problemomis, aptariamomis toliau.

4.2.1. Žanrų palyginamumas

Pateiktoji teorinė sakytinės žiniasklaidos žanrų ir požanrių klasifikacija yra tinkamiausia dabartiniam laikotarpiui – iki jo žanrai nebuvo tokie patys ir tokie aiškūs, todėl bendroje schemoje priskiriami vienai ar kitai kategorijai su išlygomis, pagal dominuojantį bruožą.

Pokalbių laidų žanras yra gana aiškus ir iš pažiūros neturėtų kelti sunkumų, tačiau sunkumų kelia skirtingas „pokalbio“ supratimas dabartiniu ir sovietiniu laikotarpiu. Kraštutiniu atveju sovietinio laikotarpio laida ar jos fragmentas, pavadinta pokalbiu, susideda iš ilgų (po kelias minutes) akivaizdžiai iš anksto parengto ir skaitomo teksto fragmentų. Vis dėlto ir tarp sovietinio laikotarpio laidų esama palyginti spontaniško, nors ir monologiškesnio nei šiandien įprasta, pokalbio pavyzdžių, kuriuos galėjo lemti ne tik laidų formatai, bet ir jose dalyvavusių žmonių asmenybės⁷.

Dabartinių TV žurnalų atitikmeniu laikomos sovietinio laikotarpio apybraižos ir dokumentiniai filmai, atitinkantys šio žanro formalius požymius (laidos vedėjo ar užkadrinio diktoriaus kalbėjimas

⁷ Pavyzdžiui, į tekstyną įtrauktos dvi 1977 m. laidos „Televizijos forumas“ (tai vieninteliai tos laidos įrašai, išlikę tik garso formatu). Nors laidos formatas tas pats, skirtingos laidos labai nevienodos dėl dalyvaujančių žmonių. Laida, kurioje dalyvauja valstybinio archyvo darbuotojai, mažiau spontaniška, joje nemažai parengto ar palyginti sukaustyto kalbėjimo ir ideologinių šampų; laida, kurioje dalyvauja meno srities dalyviai, Operos ir baletu teatro darbuotojai, įskaitant pagrindinį kalbėtoją, teatro direktorių Virgilijų Noreiką, gerokai spontaniškesnė ir gyvesnė.

derinamas su paprastų žmonių ar tų laikų įžymybių kalbėjimu). Todėl ir tekstyno aprašyme (žr. 3 lentelę 5 straipsnio skyriuje) ši žanrų kategorija, be žurnalų, apima apybraižų ir dokumentikos žanrus. Informacinis žanras palyginamumo tarp laikotarpių atžvilgiu kėlė mažiausiai sunkumų.

Atliekant kalbos raidos analizę, ypač detalesniu lygmeniu, žanrų skirtumai skirtingais laikotarpiais gali būti reikšmingi rezultatų interpretacijai. Būtent todėl pravartu kritiškai derinti kiekybinius ir kokybinius tekstyno tyrimo metodus, detalesnei analizei iš tekstyno pasirinkti struktūriškai panašiausius teksto fragmentus. Norint atskleisti bendriausius skirtumus ar panašumus, žanrų nevienodumas neturėtų būti kliūtis. Be to, panašiai įvardintų, formaliai panašių žanrų vidiniai skirtumai jau patys savaime yra reikšmingi ir rodo žiniasklaidos ir viešosios erdvės kaitą.

4.2.2. Žanrų įvairovė

Atsižvelgiant į skirtingų laikotarpių ir skirtingų žanrų paplitimo skirtingais laikotarpiais ypatumus, tekстыne taip pat teko skirti jiems nevienodą valandų skaičių.

Sovietinis laikotarpis aprašytųjų pagrindinių atrankos kriterijų atžvilgiu laikytinas sąlygiškai vienalyčiu. Nebuvo didelės laidų įvairovės, o laidų pobūdis per visą laikotarpį iš esmės nesikeitė. Pereinamajam ir ypač komerciniam laikotarpiui šiuo požiūriu būdingas daug didesnis įvairialypiškumas. Šį pokytį taip pat norėta atspindėti tekстыne, būtent todėl laiko atžvilgiu trumpesniai, palyginti su sovietiniu, dabartiniam laikotarpiui skiriamas santykinai didelis valandų skaičius, o ilgesniam sovietiniam laikotarpiui skiriama santykinai mažiau. Pereinamasis laikotarpis gana trumpas, tačiau, norint reprezentatyviai atspindėti jo metu atsirandančias kalbėjimo permainas, jam skiriamas panašus į sovietinio laikotarpio valandų skaičius (žr. 3 lentelę 5 straipsnio skyriuje).

4.2.3. Žanrų tęstinumas

Skirtingais laikotarpiais dominavo skirtingi žanrai, todėl valandų skaičius tiems patiems žanrams tarp laikotarpių tekстыne paskirstomas atitinkamai netolygiai. Tai ypač paaiškina nesubalansuotą pokalbių laidų pasiskirstymą – dabartiniu laikotarpiu joms skiriama gerokai daugiau valandų, negu sovietiniu ir pereinamuoju laikotarpiais, nes tuo metu tokių laidų buvo mažiau ir yra išlikusių nedaug jų įrašų, o dabartiniu laikotarpiu jas dėl didelio jų kiekio ir didelės vidinės paties žanro įvairovės būtina atspindėti išsamiau.

Palyginti daug valandų, ypač sovietiniu laikotarpiu, tekстыne skiriama antrosios grupės laidoms (dokumentiniai filmai ir apybraižos, žurnalai), nes tai yra didžiausią tęstinumą turintys žanrai (šių žanrų laidos egzistavo visais laikotarpiais ir jų įrašų išliko daugiausia), teikiantys bene daugiausiai galimybių diachroninei sakytinės kalbos raidos analizei. Be to, dėl technologinių priežasčių (tai yra daugkartinio rodymo laidos) šios grupės laidų daugiausiai išliko iš sovietinio laikotarpio.

5. Faktinė tekstyno struktūra ir techniniai požymiai

Atsižvelgiant į aukščiau pristatytą teorinę atrankos schemą ir kitus tekstyno sudarymą lėmusius kriterijus bei aplinkybes, sukauptas ir parengtas naudojimui tekstynas. Apibendrinta jo sudėtis valandomis parodyta 3 lentelėje, o sudėtis pagal žodžių skaičių – 4 lentelėje.

Žanrų grupė	Sovietinis laikotarpis (1960–1987)	Pereinamasis laikotarpis (1988–1992)	Dabartinis laikotarpis (1993–2011)	Iš viso
Pokalbių laidos (televizija/radijas)	6 (5/1)	5,5 (5,5/0)	18,5 (16/2,5)	30
Dokumentiniai filmai ir apybraižos, žurnalai (televizija/radijas)	9 (7,5/1,5)	4,5 (4,5/0)	7 (7/0)	20,5
Informacinės laidos (televizija/radijas)	6 (1/5)	2 (2/0)	4,5 (2,5/2)	12,5
Iš viso (televizija/radijas)	21 (13,5/7,5)	12 (12/0)	30 (25,5/4,5)	63

3 lentelė. Sakytinės žiniasklaidos tekstyno sudėtis valandomis

	Sovietinis laikotarpis (1960–1987)	Pereinamasis laikotarpis (1988–1992)	Dabartinis laikotarpis (1993–2011)	Iš viso
Žodžių skaičius	98 000	65 000	195 000	358 000

4 lentelė. Sakytinės žiniasklaidos tekstyno sudėtis pagal žodžių skaičių

Savaime suprantama, kad sovietinius metus reprezentuoja laidos iš vienintelio tuo metu egzistavusio transliuotojo – valstybinio Lietuvos televizijos ir radijo; komercinis laikotarpis, be Lietuvos nacionalinio radijo ir televizijos, įtraukia ir pagrindinius komercinius televizijos kanalus: TV3, LNK, Baltijos TV bei dabartiniu laikotarpiu svarbiausias radijo stotis M-1 ir Radiocentrą.

Sakytinės žiniasklaidos tekстыne šiandien sukaupia 63 val. transkribuotų vaizdo ir garso įrašų (lentelėje pateikiami suapvalinti skaičiai). Tai yra apie 140 laidų ar jų fragmentų⁸. Tekstyno įrašų trukmė skaičiuojama pagal įrašų trukmę. Ilgos reklaminės pauzės, muzikinės užsklandos ar kiti nekalbiniai intarpai buvo iškirpti, išrašė apie tai nurodant komentaru. Vis dėlto laidose yra likusių nedidelių įvairios trukmės nekalbinių intarpų, todėl grynoji kalbėjimo trukmė tekстыne yra kiek mažesnė. Atskirų įrašų trukmė labai įvairuoja: kai kurie įrašai apima tik nedidelį išlikusį laidos fragmentą (šiuo metu tekстыne trumpiausias įrašas yra radijo laidos apie Vinco Mykolaičio-Putino jubiliejų 1963 m. fragmentas, trunkantis 2 min. ir 30 sek., trumpiausia atskira laikytina laida – 1998 m. radijo stoties M-1 eteryje skaitomų žinių įrašas, trunkantis 1 min. 30 sek.), dauguma – visą laidą ar tam tikrą logiškai atskirą ilgos laidos dalį (ilgiausia laida – „Kūrybos vakaras“ su Donatu ir Raimundu Banioniais, su intarpais trukęs 1 val. ir 16 min).

Tekstyno žodžių skaičius – netoli 360 tūkstančių. Lyginant su kitais lietuviškais ir ypač užsienio tekstynais, jis nėra didelis. Pavyzdžiui, VDU su partneriais parengtame Sakytinės kalbos tekстыne sukaupia 230 val. įrašų, išrašai sudaro 680 tūkst. žodžių⁹. Vis dėlto būdamas tikslinis, specializuotas, Sakytinės žiniasklaidos tekstynas laikytinas palyginti nemažu (reikia turėti galvoje ir sunkumus, kilusius ieškant archyvinės medžiagos). Tekstynas rengtas kaip statinis, užbaigtas darbas, tačiau esant galimybėms bus pildomas, išlaikant tuos pačius sandaros principus.

Iš viso į tekстыną įtrauktose laidose dalyvauja per 1000 kalbėtojų; kai kurie kalbėtojai po keletą kartų ir neretai skirtingais vaidmenimis. Suprantama, kalbėtojų pasisakymai yra labai nevienodos trukmės, nuo kelių žodžių iki didelės dalies laidos. Rengiant laidų išrašus, visi kalbėtojai buvo žymimi unikaliais kodais ir priskiriami vienam iš 6 išskirtų vaidmenų tipų: laidos vedėjas, pranešėjas/diktorius/balsas už kadro, įžymybė, ekspertas, „herojus“ (paprastas žmogus, kurio veiksmai ar gyvenimas pristatomi laidoje), *vox populi* (paprastas žmogus, paprašytas trumpai išsakyti savo nuomonę užduotu klausimu). Taip pat buvo pažymima, jei veikėjas kalba telefonu, nors tokių atvejų buvo labai mažai. Didžiausią dalį tekстыne sudaro ekspertai (netoli 30 proc.), taip pat „herojaus“ tipo kalbėtojai (apie 25 proc.), mažiau yra pranešėjų (netoli 20 proc.), *vox populi* vaidmenų kalbėtojų (netoli 15 proc.) ir įžymybių (apie 10 proc.), mažiausiai – laidų vedėjų (netoli 5 proc.). Kadangi nebuvo siekiama tekstyno balanso veikėjų atžvilgiu, tai nelaikytina trūkumu ir

⁸ Nurodomas apytikslis skaičius, nes kai kuriais atvejais nėra aiškūs laidų atskyrimo kriterijai arba trūksta archyvinio aprašymo duomenų.

⁹ Krinta į akis nevienodos laidų įrašų ir išrašytų žodžių skaičiaus proporcijos Sakytinės lietuvių kalbos tekстыne ir Sakytinės žiniasklaidos tekстыne: atitinkamai 230 val./680 tūkst. ir 63 val./360 tūkst. Morfologinį kodavimą atlikusios ekspertės pastebėjimu, žiniasklaidos tekstai iš tiesų buvo gerokai „tankesni“, kalbėtojai kalba greičiau, yra mažiau pauzių. Veikiausiai tai susiję ir su rašytinių tekstų skaitymu.

laikoma, kad veikėjai atspindi viešosios erdvės ypatumus apskritai ir skirtingais laikotarpiais (daugiau žr. Nevinskaitė 2012). Taip pat buvo bandoma fiksuoti kalbėtojų amžių, tačiau dėl jo nustatymo sunkumų, ypač šifruojant garso įrašus be vaizdo, ši informacija nėra išsami ir nevisiškai patikima.

Tekstynas parengtas naudojantis CHILDES projekte sukurta CLAN programine įranga, laikantis CHILDES projekto išrašymo ir kodavimo reikalavimų. Išrašytas tekstas laiko žymomis yra susietas su vaizdu ir garsu. Tekstynas jau jo kūrimo metu buvo koduojami kai kurie kalbos bruožai (adresato įvardijimai, juokas, pauzės, kai kurios sintaksinės konstrukcijos, šnekamosios leksikos bruožai ir kt.); tekstynas yra morfologiškai anotuotas. Tekstyno išrašymo ir pirminio kodavimo darbus atliko filologijos ir kitų humanitarinių krypties studentai¹⁰; kodai tikrinti ir tikslinti, papildomi kalbos bruožai (pvz., fonetikos) koduoti pačių projekte dalyvavusių tyrėjų.

6. Baigiamosios pastabos

Didelės apimties sakytinės žiniasklaidos kalbos tyrimai neįmanomi be reprezentatyvių, sistemingai ir apgalvotai parengtų tekstynų. Tikimasi, kad projekte „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“ parengtas tekstynas paskatins platesnius sakytinės viešosios kalbos tyrimus. Tekstynas buvo kuriamas minėto projekto tikslams, tačiau, be jokios abejonės, sukaupta medžiaga gali būti panaudojama ir keliant kitus tyrimo klausimus bei kitų sričių tyrėjų.

Nors dėl riboto įrašų prieinamumo tekstynas negali būti laikomas visiškai reprezentatyviu ir subalansuotu, atsirenkant iš tekstyno pagal formalius požymius ekvivalentiškas laidas ar jų fragmentus ir kitaip atsižvelgiant į aptartus tekstyno apribojimus, be abejonės gali būti atliekama lyginamoji analizė ir daromos išvados apie kalbos kaitą. Atlikti pirmieji tyrimai rodo, kad tekstyno apimtis leidžia daryti apibendrinimus apie viešosios kalbos kaitą skirtingais laikotarpiais ir pateikti logiškai įtikinamas, taigi, tikėtina, pakankamai reprezentatyvias, išvadas (žr. Aleksandravičiūtė ir Vaicekauskienė 2012; Girčienė 2013; Čičirkaitė 2012).

Apžvelgtos įrašų prieinamumo problemos, su kuriomis susidūrė tekstyno kūrėjai, kelia mintį, kad tyrėjai turėtų jau dabar pradėti kaupti dabartinės viešosios kalbos duomenis palyginamiesiems retrospektyviems tyrimams ateityje. Pasinaudojant vaizdo įrašymo technologijomis arba laidų transliacijomis internete būtų galima kaupti asmeninius laidų archyvus – norint juos naudoti

¹⁰ Dėkojame daugiausiai prisidėjusiems išrašinėtojams: Emilijai Čekuolytei, Simonai Ševelytei, Inai Usaitei ir Mykolui Zovei.

tyrimams, reikėtų išspręsti autorinių teisių klausimus, bet bent jau būtų išvengiama didelio darbo ieškant šių įrašų ir sprendžiant technologines problemas. Šiuo metu audiovizualinės žiniasklaidos transliacijoms ir saugojimui naudojamos skaitmeninės technologijos iš pažiūros leidžia lengvesnį įrašų saugojimą, tačiau niekas nėra garantuotas ir dėl jų tęstinumo, tuo labiau, kad, kaip rodo apžvelgta patirtis, radijas ir televizija, ypač komerciniai, ne visada suinteresuoti jų ilgalaikiu išsaugojimu.

Šiuo metu tekstynas prieinamas projekto dalyviams, tačiau ateityje, išsprendus autorinių teisių klausimus, jis galėtų tapti prieinamas platesniam tyrėjų ratui. Jei bus galimybių, ateityje planuojama parengti viešai internete prieinamą jo versiją.

Literatūra

- Aleksandravičiūtė, S. ir L. Vaicekauskienė. 2012. Juokai juokais, bet šiandien juokiamės daugiau ir kitaip: juokas kaip TV ir radijo diskurso bruožas 1960–2010 metais. *Darbai ir dienos* 58, 283–295.
- Baker, P., A. Hardy ir T. McEnery. 2006. *A Glossary of Corpus Linguistics*. Edinburgh, Edinburgh University Press.
- Biber, D. Representativeness in Corpus Design. *Literary and Linguistic Computing* 8 (4), 243–257.
- Bignell, J. 2004. *An Introduction to Television Studies*. London: Routledge.
- Čičirkaitė, R. 2012. Kirčiuotų trumpųjų balsių ilginimo polinkiai 1960–2011 m. informacinėse laidose. *Kalbos kultūra* 85 [įteikta spaudai].
- Dabašinskienė, I. ir L. Kamandulytė. 2009. Corpora of spoken Lithuanian. *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian papers in applied linguistics* 5, 67–77.
- Fairclough, N. 1995. *Media Discourse*. London: Edward Arnold.
- Girčienė, J. 2009. Daugiskaitinių ir vienaskaitinių kreipimosi formų konkurencija viešojoje sakytinėje vartosenoje (sociolingvistinis aspektas). *Kalbos kultūra* 82, 214–234.
- Girčienė, J. 2010. Nominatyviniai adresato įvardijimai visuomeninius klausimus nagrinėjančiose pokalbių laidose: sociolingvistinis aspektas. *Žmogus kalbos erdvėje* 6, 690–698.
- Girčienė, J. 2013. Adresato įvardijimo raida. *Žmogus kalbos erdvėje* 7, 613–621.
- Girčienė, J., L. Nevinskaitė ir G. Tamaševičius. 2010. *Sakytinių viešosios kalbos tekstų atrankos aprašas*. Rengta projekto „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“ 2010 m. ataskaitai [nepublikuotas rankraštis].
- Girčienė, J., ir G. Tamaševičius. 2012. Five Decades of Television: from Language Homophony to Polyphony. *Lituanus. The Lithuanian Quarterly Journal of Arts and Sciences* 58 (2), 31–43.
- Hedin, T. 2005. *Changing Identities. Language Variation on Czech Television*. Stockholm: Stockholm University.
- Hunston, S. 2008. *Corpus Linguistics: An International Handbook. Vol. 1*. A. Lüdeling and M. Kytö (eds.). Berlin: Walter de Gruyter. 154–168.

- Kenedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Lüders, K., L.Prøitz ir T. Rasmussen. 2010. Emerging Personal Media Genres. *New Media & Society*, 12 (6), 947–963.
- Marcinkevičienė, R. 2000. Tekstynų lingvistika: teorija ir praktika. *Darbai ir dienos* 24, 7–64.
- McEnery, T. ir A. Wilson. 1996. *Corpus Linguistics*. Edinburgh: University Press.
- Nevinskaitė, L. 2006. *Viešosios erdvės transformacija Lietuvoje 1988–2000 m.: laikraščių atvejis*. Daktaro disertacija [nepublikuotas rankraštis]. Vilnius.
- Nevinskaitė, L. 2012. On the public sphere and its participants. *Lituanus. The Lithuanian Quarterly Journal of Arts and Sciences*, 58 (2), 44–57.
- Pečiulis, Ž. 2000. Televizijos programų plėtra. *Naujosios žiniasklaidos formavimasis Lietuvoje (1988–1998 m.)*. Vilnius: Žurnalistikos institutas, Lietuvos žurnalistų sąjunga. 232–254.
- Pečiulis, Ž. 2007. *Iki ir po televizijos. Žvilgsnis į XX amžiaus audiovizualinės masinės komunikacijos fenomeną*. Kaunas: Versus Aureus.
- Summers D. 1991. *Longman/Lancaster English Language Corpus: Criteria and Design*. Harlow: Longman.
- Tamaševičius, G. 2009. Kalbos paprastėjimo polinkiai viešojoje erdvėje. *Kalbos kultūra* 82, 138–148.
- Tamaševičius, G. 2010. „Blogo“ kalba ir sociolingvistinė kompetencija. *Darbai ir dienos* 53, 189–198.

Retrospektyvus Lietuvos sakytinės žiniasklaidos tekstynas: rengimo metodologija ir patirtis

Laima Nevinskaitė

Santrauka

Straipsnio tikslas – pristatyti akademinėi bendruomenei projekte „Lietuvių kalba: idealai, ideologijos ir tapatybės lūžiai“ parengtą apie 65 val. sakytinės žiniasklaidos kalbos 1960–2010 m. tekstyną, aptarti metodologinius jo kūrimo principus ir problemas, pasidalyti praktine tekstų paieškos ir atrankos patirtimi.

Vienas svarbiausių reikalavimų tekstynui yra jo reprezentatyvumas ir subalansuotumas. Galimas būdas jų siekti yra išskirti objektyviai apibrėžiamus tekstų tipus ir pagal juos konstruoti tekstynų sandarą. Šio tekstyno sandara buvo modeliuojama pagal du kriterijus: sakytinės žiniasklaidos raidos laikotarpius ir žanrus. Išskiriami trys laikotarpiai: sovietinis 1960–1987 m., pereinamasis 1988–1992 m. ir dabartinis nuo 1993 m. bei trys žanrų grupės: pokalbių laidos (skaidomos į pokalbių-interviu, debatų ir pokalbių šou tipus); dokumentiniai filmai ir apybraižos, žurnalai ir informacinės laidos. Pagal šiuos kriterijus ir siekiama sudaryti subalansuotą tekstyną. Straipsnyje aptariami sunkumai, iškilę įgyvendinant tekstyną pagal šią schemą: medžiagos prieinamumo sunkumai dėl

technologijų įtakos ir archyvinių šaltinių organizacinių veiksnių, medžiagos balanso tarp laikotarpių klausimas, žanrų palyginamumo tarp skirtingų laikotarpių, nevienodos žanrų įvairovės skirtingais laikotarpiais ir žanrų tęstinumo problemos.

Straipsnio pabaigoje pristatoma faktinė tekstyno struktūra ir dydis (63 val., apie 350 tūkst. žodžių) ir konstatuojama, kad nors dėl įrašų prieinamumo ir kitų sunkumų tekstynas nėra visiškai reprezentatyvus ir subalansuotas, jis yra pakankamas, kad būtų sėkmingai panaudojamas tyrimams; tai patvirtina pirmųjų tyrimų iš jo medžiagos rezultatai. Techniniu požiūriu tekstynas atitinka įprastus tekstynams keliamus reikalavimus: jis išrašytas pasitelkus plačiai naudojamą CHILDES projekto CLAN programinę įrangą; išrašytas, sukoduotas (koduoti kai kurie diskurso, leksikos ir sintaksės bruožai) ir morfologiškai anotuotas pagal CHILDES projekto konvencijas; unikaliais kodais fiksuojami kalbėtojai; išrašytas tekstas yra susietas su vaizdu/garsu.

Methodology and Experience of Building the Retrospective Corpus of Lithuanian Broadcast Media

Laima Nevinskaitė

Summary

The article aims to introduce the corpus of broadcast media (radio and television), which was compiled in the framework of the project *Lithuanian Language: Ideals, Ideologies and Identity Shifts* to the academic community. The corpus includes about 63 hours of transcribed recordings from 1960 to 2010. The article discusses theoretical principles of corpus sampling, which were based on the criteria of time periods and genres; the methodological issues encountered when constructing the sampling scheme; shares the practical experience of selecting and gathering recordings to be included into the corpus; and presents the actual structure of the corpus.

One of the main requirements for any corpus is its representativeness and balance. One possible way to achieve them is to distinguish objectively defined text types and to build the corpus along these lines. Thus the composition of the corpus of the broadcast media was designed on the basis of two criteria: periods of broadcast media development and genre. There are three periods distinguished: Soviet 1960–1987, transitional 1988–1992 and contemporary 1993–now. In respect of genre, three groups of programs are included: talk programs (further subdivided into the types of interview, debate and talk-show); documentaries, features and journal programs; information programs. The article discusses problems that were encountered when trying to implement the corpus along these lines: problems of availability of materials due to technological peculiarities of

different periods and organisational factors of archive institutions; the issue of balance between the periods; the problems of genre comparability and different extent of diversity of genres in different periods, and continuity of genres.

Finally, the composition and the size of the corpus are presented (63 hours of recordings, about 350 thousand words). The paper concludes that despite the limited availability of materials and other problems discussed above which is why the corpus cannot be regarded as perfectly representative and balanced, it is sufficient for research into public language change. This was confirmed by tentative research studies done on its basis. The corpus meets the usual technical requirements: the transcriptions have been made in CLAN software developed within the CHILDES project, the recordings have been transcribed, coded and morphologically annotated following the conventions of the CHILDES project, the speakers have been assigned individual codes, and the transcriptions have been linked to the sound/image files.

Įteikta: 2013 m. liepos mėn.