

SANTRAUKA. Straipsnyje pateikiama Lietuvos Respublikos Seimo narių kalbinės raiškos parlamentinių debatų metu analizė, atsižvelgiant į jų politinę orientaciją – kairę, centrą ir dešinę. Tyrime taikyti kompiuterinės stilometrijos metodai leidžia apibūdinti ir paaiškinti ryšius tarp psichologinių bei sociologinių autoriaus charakteristikų ir išanalizuoti autoriaus stilių. Tyrime politinė orientacija traktuojama kaip tiriamųjų parlamento narių savybė, t. y. žvelgiama iš stiliaus perspektyvos: politinės kairės, politinio centro ir politinės dešinės stilius. Tyrimo rezultatai parodė, jog kalbinės raiškos, žyminčios politinę orientaciją, apimtis labai skiriasi ir daro poveikį parlamentinių debatų metu vartojamam žodynui.

RAKTAŽODŽIAI: Seimo posėdžių stenogramos, stilometrija, kompiuterinė stilistika, politinė orientacija.

ĮVADINĖS PASTABOS

Tradiciškai politinių partijų priskyrimas tam tikrai politinei orientacijai priklauso nuo jų „emocinio prisirišimo“ ir artumo socialinėms grupėms, susijusioms su šiomis partijomis (Green *et al.*, 2002). Pagrindinės politinės vertybės atitinka abstrakčius, norminamuosius įsitikinimus apie žmoniją, visuomenę ir viešuosius reikalus (Kinder, 1998). Kalbant konkrečiau, politines partijas į politinį spektrą galima išdėlioti pagal jų požiūrį į lygias galimybes, valdžios įtakos apimtį, tradicines šeimoms vertybes, moralinį pakantumą, socialines problemas, ekonominę politiką etc. (Goren, 2005; Morgan, *et al.*, 2010). Kadangi partijoms rūpi skirtingi su jomis susijusių socialinių grupių interesai arba atskiri šių interesų aspektai, galima daryti prielaidą, jog partijos, priklausančios nevienodai politinei orientacijai, savo požiūrį išreiškia skirtingai.

Laikantis aukščiau iškeltos prielaidos, šiame straipsnyje pristatoma Lietuvos Respublikos parlamento narių kalbinės raiškos parlamentinių debatų metu analizė, atsižvelgiant į jų politinę orientaciją – kairę, centrą ir dešinę. Tyrime naudoti

kompiuterinės stilometrijos metodai leidžia apibūdinti ir paaiškinti ryšius tarp psichologinių bei sociologinių autoriaus charakteristikų ir išanalizuoti autoriaus stilių. Taigi šiame tyrime politinė orientacija traktuojama kaip tiriamųjų parlamento narių savybė, t. y. žvelgiama iš stiliaus perspektyvos: politinės kairės stilius, politinio centro stilius ir politinės dešinės stilius.

Žvelgiant į stiliaus ir stilistikos tyrimų istoriją, Lietuvoje stilistinių tyrimų pradžia sietina su Juozo Pikčilingio apibrėžtos autoriaus individualaus stiliaus sąvokos atsiradimu (Pikčilingis, 1971). Nuo tada atlikta nemažai aprašomojo pobūdžio kalbinių tyrimų, kuriuose dėmesys skirtas autoriaus stiliaus (idiolekto) ypatybėms. Tokių tyrimų yra atlikusi, pvz., Anelė Žalkauskienė (2000), analizuojant tekstus teismų ekspertizėms siūlanti atsižvelgti į vartojamus sinonimus. Gintarė Žalkauskaitė elektroninių laiškų autorystės nustatymą sieja su konkrečiam autoriui būdinga ir nebūdinga raiška: vartojamus žodžius, skyrybos ir grafinius ženklus etc. (Žalkauskaitė, 2014). Kalbant apie atskirų socialinių charakteristikų raišką tekste, pvz., vyrų ir moterų kalbėjimo skirtumus, pasakytina, kad tai analizuota atliekant dažniausiai kokybinę lingvistinę analizę (Galinskaitė, 2014; Girčienė, 2009; Savickienė, 2006).

Stilometrijos metodai Lietuvoje dar tik pradedami taikyti. Pirmieji automatiniai stiliaus analizės, autorystės ir autorių charakteristikų nustatymo tyrimai atlikti naudojant Lietuvos Respublikos Seimo posėdžių stenogramas (Mandravickaitė, 2013; Kapočiūtė-Dzikienė ir Krupavičius, 2014; Kapočiūtė-Dzikienė *et al.*, 2014a; Kapočiūtė-Dzikienė *et al.*, 2014b; Šarkutė ir Utka, 2015). Šiems tyrimams taikyti skirtingi metodai (prižiūrimi ir neprižiūrimi mašininio mokymo metodai), skirtingi požymiai (dažniausi žodžiai, tarnybiniai / funkciniai žodžiai, simbolių n-gramos, žodžių n-gramos, lemos ir jų n-gramos, kalbos dalių žymelės etc.). Jurgita Kapočiūtė-Dzikienė, Ligita Šarkutė ir Andrius Utka (2014c) sprendė Lietuvos Respublikos Seimo stenogramų automatinio priskyrimo politinei orientacijai (politinė kairė-politinis centras-politinė dešinė) uždavinį, tačiau jų tyrimo tikslas buvo rasti tinkamiausią metodą ir požymius teisingam stenogramų suklasifikavimui pagal jau minėtą politinę orientaciją, pačią kalbinės raiškos ir skiriamųjų požymių analizę paliekant kitiems tyrėjams.

Pasaulyje automatinius autorių stilių tyrimus pradėjo Frederickas Mostellier ir Davidas Wallace'as, ginčytinos autorystės tekstus iš „The Federalist Papers“ rinkinio, skatinančio ratifikuoti JAV Konstituciją, priskyre Džeimsui Medisonui (Mosteller and Wallace, 1964). Vėliau stilometriją imta taikyti ne tik autorystės priskyrimui, bet ir kitus uždavinius sprendžiant, pvz., įvairių teksto autoriaus charakteristikų nustatymui, kitaip – autorių profiliavimui.

Stilometrijos metodai autorių profiliavimui naudoti norint nustatyti tekstų autorių lytį (Argamon *et al.*, 2003; Argamon *et al.*, 2007; 2009; Koppel *et al.*,

2002; Sarawgi *et al.*, 2011), amžių (Dahllöf, 2012), psichologinį asmenybės tipą (Luyckx and Daelemans, 2008; Celli, 2012), gimtąją kalbą (kai autoriaus anglų kalba nėra gimtoji) (Bykh and Meurers, 2012). Matsas Dahllöfas (2012) atliko eksperimentus automatiniam politinės priklausomybės priskyrimui. Kai kurie autoriai savo tyrimais siekė nustatyti ne atskiras savybes, o sukurti profilį, apimantį daugiau nei vieną charakteristiką (Brennan, 2012; De-Arteaga *et al.*, 2013; Mechti *et al.*, 2013; Rustagi *et al.*, 2009). Eksperimentai atlikti su skirtingoms sritims priklausančiais tekstais, pvz., tinklaraščiais (Gianfortoni *et al.*, 2011), elektroniniais laiškais (Cheng *et al.*, 2009), interneto forumo įrašais (Nguyen *et al.*, 2011), rašiniiais (Luyckx and Daelemans, 2008).

Kita vertus, stilometrija vis dar dažniausiai taikoma baigtinio rinkinio (angl. *closed set*) uždavinių sprendimui – kai nežinomas tekstas priklauso vienam iš keleto žinomų autorių. Šio tipo uždavinius sprendė ir Kapočiūtė-Dzikiienė su kolegomis (Kapočiūtė-Dzikiienė *et al.*, 2014a; Kapočiūtė-Dzikiienė *et al.*, 2014b) bei Šarkutė ir Utkā (2015). Realūs uždaviniai, ypač teisminės ekspertizės srityje, paprastai patenka tarp atviro rinkinio (angl. *open set*) (Stamatatos, 2009) uždavinių. Atviro rinkinio uždavinys yra tuomet, kai tiriamųjų autorių, siekiant nustatyti nežinomo teksto autorystę, yra labai daug, bet tikrasis nežinomo teksto autorius nebūtinai yra tarp šių tiriamųjų autorių, pvz., tiriant interneto komentarų autorystę. Šiame straipsnyje kalbinės raiškos ir skiriamųjų požymių Lietuvos Respublikos Seimo posėdžių stenogramose, atsižvelgiant į priskyrimą skirtingai politinei orientacijai, analizė bus atliekama laikantis atviro rinkinio uždavinio perspektyvos.

LIETUVOS RESPUBLIKOS SEIMO STENOGRAMŲ TEKSTYNAS

Tyrimui naudojome Lietuvos Respublikos Seimo stenogramų tekstyną, parengtą vykdant projektą „Automatiniai autorių ir autorių grupių individualaus stiliaus nustatymo tyrimai“ (ASTRA) (Nr. LIT-8-69)¹. Tekstyną sudaro Lietuvos Respublikos Seimo posėdžių stenogramos nuo 1990 m. kovo mėnesio iki 2013 m. gruodžio mėnesio imtinai. 28 379 pavyzdžiai priklauso Seimo nariams, priklausantiems politinio centro partijoms, 42 233 – politinės dešinės partijū nariams ir 40 296 – politinės kairės partijū nariams. Iš viso tekstyną sudaro beveik 24 mln. žodžių, o kiekvieną teksto pavyzdį – ne mažiau nei 100 žodžių; į tekstyną pateko tik parlamentarai, turintys ne mažiau nei 200 kalbų / pasisakymų (Kapočiūtė-Dzikiienė *et al.*, 2014a).

Tyrimams visi pavyzdžiai buvo sujungti į tris didelius dokumentus atsižvelgiant į politinę orientaciją. Tam, kad tekstų apdorojimas vyktų sparčiau, šie trys dideli

¹ Prieiga internete: <<http://dangus.vdu.lt/~jkd/>>.

dokumentai buvo suskaidyti į mažesnius vienodos apimties dokumentus. Vienai daliai tyrimų kiekvienas didysis dokumentas buvo suskaidytas į 20 mažesnių dokumentų, kitai – į 10. Taip politinei kairei, centrui ir dešinei buvo priskirta po 20 arba po 10 vienodos apimties pavyzdžių.

TYRIMO METODAS

Kompiuterinės stilometrijos metodai leidžia apibūdinti ir paaiškinti priežastinius ryšius tarp psichologinių bei sociologinių autoriaus charakteristikų, išanalizuoti autoriaus stilių. Stilometrija remiasi hipoteze, jog individo stilius yra unikalus. Ši hipotezė vadinama stilomo hipoteze (angl. *human stylome hypothesis*) (Halteren *et al.*, 2005). Žvelgiant iš informacijos paieškos (angl. *information retrieval*) perspektyvos, stilometrija leidžia įgyti metažinių: ką galima iš teksto suprasti apie jo autorių (Daelemans, 2013). Šios metažinios apie autorių gali būti lytis, amžius, psichologinės ypatybės (pvz., introvertiškumas), politinės priklausomybė etc.

Visas panašias savybes galima nustatyti automatiškai, naudojant stilometrijos metodus. Patikimas automatinis lyties, amžiaus, psichologinių ypatybių ir pan. nustatymas arba profiliavimas bei autorystės priskyrimas naudingas daugeliui sričių: literatūroje, teismo lingvistikoje, rinkodaroje (pvz., siekiama nustatyti konkrečiu produktu besidominčių vartotojų charakteristikas), kalbos psichologijoje, sociolingvistikoje, medicinos diagnostikoje (pvz., nustatant Alzheimerio ligą ir sekant jos progresavimą (Garrard, 2009)), aptinkant grėsmes internete (pvz., identifikuojant pedofilus (Rashid *et al.*, 2009)) etc.

Be žmogaus stilomo hipotezės, stilometrija remiasi prielaida, jog stilius gali būti „išmatuotas“ statistiškai apdorojant įvairias jo charakteristikas. Pagrindinis stilometrijos ir įprastesnio teminio klasifikavimo skirtumas yra tai, kad stilometrinei analizei naudojamos dažniausios, paprastai nuo turinio nepriklausančios teksto charakteristikos (leksikos, morfologinės, sintaksinės, simbolių etc.).

Stilometrijos metodų taikymui Lietuvos Respublikos Seimo stenogramų kalbinės raiškos, atsižvelgiant į politinę orientaciją, analizei remtasi Maciejaus Ederio, Jano Rybicki ir Mike'o Kestemonto siūloma analizės eiga (Eder, 2010; Eder ir Rybicki, 2013; Kestemont, 2014; Rybicki ir Eder, 2011), naudojant programinę paketą *Stylo* (Eder *et al.*, 2014), skirtą atlikti stilometrines analizes su matematinės statistikos programa R.

Aukščiau minėtos stilometrines analizės elementai apima analizuojamus tekstus, juos skiriančiuosius požymius (pvz., dažniausius žodžius, simbolių n-gramas, žodžių n-gramas, lemų n-gramas, kalbos dalių žymeles, sintaksinės analizės

žymeles etc.; tinkamiausi požymiai ir jų skaičius analizei parenkami eksperimentiškai, kadangi nėra universalaus sprendimo), atstumo matas (atstumui arba distancijai tarp tekstų išmatuoti – kuris kuriam artimesnis ir kartu panašesnis; žr. toliau) ir hierarchinę klasterinę analizę (taip pat ir daugiamačių skalių analizę, pagrindinių komponenčių analizę etc.). Hierarchinės klasterinės analizės rezultatų vizualizavimui naudojamos dendrogramos išdėsto teksto pavyzdžius pagal panašumą, t. y. kuo tekstai (jų grupės) panašesni, tuo jie arčiau vienas kito, o prisijungimo prie šakos pozicija (tolstant nuo objektų) rodo augantį skirtumą tarp jų ar jų grupių (klasterių).

Kalbant apie skiriamuosius požymius, dažniausi žodžiai, tarnybiniai / funkciniai žodžiai (daugeliu atveju jie yra dažnių lentelės viršuje (Hochmann *et al.*, 2010; Sigurd *et al.*, 2004)) yra patys populiariausi atliekant stilometrinę analizę (Burrows, 1992; Hoover, 2007; Eder, 2013b; Eder and Rybicki, 2013). Tarnybiniais / funkciniais žodžiais vadinami pagal „Kalbotyros terminų žodyną“ (Gaivenis ir Keinys, 1990) – prielinksniai, dalelytės ir jungtukai. Sekant anglų kalbos tradicija, tarnybinių žodžių sąrašas papildytas įvardžiais ir jaustukais (Bell *et al.*, 2009). Dažniausi žodžiai (angl. *Most Frequent Words, MFW*) kaip skiriamieji požymiai (paprasčiausias pasirinkimas nei tarnybiniai / funkciniai žodžiai, nes nereikia turėti iš anksto parengto jų sąrašo ar jų išsirinkti iš analizuojamų tekstų) yra laikomi neutraliais teminiu atžvilgiu, ir tyrimai, atlikti juos naudojant, buvo gana sėkmingi (Juola and Baayen, 2005; Holmes *et al.*, 2001; Burrows, 2002).

Lietuvos Respublikos Seimo stenogramų kalbinės raiškos stilometrinei analizei, atsižvelgiant į politinę orientaciją, taip pat naudoti dažniausi žodžiai. Be to, darant prielaidą, kad parlamentarai profesinėje aplinkoje (šiuo atveju Lietuvos Respublikos Seime) kalba gana formalizuota kalba, vartoja specifinius posakius, žodžių junginius, analizei naudotos ir dažniausios žodžių bigramos. Šį pasirinkimą galima pagrįsti dar ir tuo, kad stereotipinė kalba (angl. *formulaic language*) kaip skiriamasis požymis leidžia detalesnę rezultatų interpretaciją nei naudojant tik dažniausius žodžius (Antonia *et al.*, 2014).

Kaip jau minėta, tyrimai atlikti su *Stylo* paketu, skirtu stilometrinei analizei su matematinės statistikos programa R (Eder *et al.*, 2014). Detalizuojant stilometrines analizės procesą, visų pirma, imant visus analizuojamų tekstų pavyzdžius, suformuojamas dažninis analizuojamų tekstų žodžių (ar kitų analizei naudojamų skiriamųjų požymių) sąrašas, vėliau normalizuojamas Z-įverčiais (kitai – standartiniais įverčiais). Z-įverčiai apskaičiuojami atimant tam tikro požymio vidutinį dažnį viename teksto pavyzdyje iš jo vidutinio dažnio visuose analizuojamų tekstų pavyzdžiuose kartu paėmus ir šį skirtumą padalijus iš standartinio nuokrypio (Hoover, 2004). Tuomet naudojant Burrowso deltos (1) matą (arba jo variantą)

(Burrows, 2002) randama dviejų tekstų nepanašumo skaitinė vertė, kuri yra Z -įverčių skirtumų visiems tiriamiems skiriamiesiems požymiams vidurkis tarp šių dviejų tekstų:

$$\Delta_{(AB)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{f_i(A) - f_i(B)}{\sigma_i} \right|$$

(1)

kur n – dažniausių žodžių skaičius, f_i – žodžio i dažnis, A ir B – lyginami tekstų pavyzdžiai ir σ_i – žodžio i dažnio standartinis nuokrypis.

Toliau, naudojant visas porines nepanašumo vertes tarp tekstų, suformuojama atstumų arba distancijų matrica. Galiausiai naudojamas hierarchinis klasterizavimas, kurio dendrogramos leidžia pamatyti tekstų pavyzdžių artimumą / panašumą, juos lyginant vieną su kitu.

Minėtoji Burrowso delta yra turbūt pats populiariausias atstumo matas stilometri-
nėje analizėje (Burrows, 2002; Rybicki and Eder, 2011). Verta paminėti, jog rezultatų, naudojant šį matą, tikslumas priklauso nuo Z -įverčių, analizuojamų tekstų skaičiaus ir balanso tarp jų atsižvelgiant į teksto pavyzdžių kiekį, ilgį bei autorių skaičių (Stamatatos, 2009). Nors Burrowso delta buvo sėkmingai naudota angliškiems ir vokiškiems tekstams, rezultatai su labiau linksniuojamų kalbų tekstais (lotynų, lenkų) buvo ne tokie sėkmingi (Rybicki and Eder, 2011). Tai lėmė, jog Lietuvos Respublikos Seimo stenogramų stilometri-
nėi analizei, siekiant išanalizuoti parlamentarų kalbinę raišką atsižvelgiant į partijų, kurioms jie priklauso, politinę orientaciją, buvo pasirinktas deltos variantas. Ederio delta (2) yra modifikuota klasikinė Burrowso delta, tačiau ji suteikia daugiau svorio dažniams analizuojamiems požymiams ir atitinkamai „permatuoja“ retesnius analizuojamus požy-
mius, siekiant išvengti atsitiktinių retų požymių atliekamoje analizėje (Eder *et al.*, 2014). Šis matas buvo sukurtas naudoti atliekant stilometrines analizes su itin linksniuojamų kalbų (pvz., lietuvių) tekstais. Vis dėlto ir ši delta, kaip ir visos kitos, išlaiko jautrumą analizuojamų pavyzdžių skaičiui. Ederio delta skaičiuojama pagal šią formulę:

$$\Delta_{(AB)} = \frac{1}{n} \sum_{i=1}^n \left(\left| \frac{f_i(A) - f_i(B)}{\sigma_i} \right| * \frac{n - n_i + 1}{n} \right)$$

(2)

kur n – dažniausių žodžių skaičius, f_i – žodžio i dažnis, A ir B – lyginami tekstų pavyzdžiai ir σ_i – žodžio i dažnio standartinis nuokrypis.

Nors hierarchinis klasterizavimas, pasirinktas tekstinių pavyzdžių atvaizdavimui atsižvelgiant į jų tarpusavio ryšius (panašumą–nepanašumą) pagal politinę orientaciją, yra jautrus skiriamųjų požymių skaičiui arba pavyzdžių grupavimo metodo pasirinkimui (Eder, 2013a), šiame tyrime gauti rezultatai buvo gana stabilūs. Siekiant patikrinti hierarchinio klasterizavimo patikimumą, pasitelkta kilpos² (angl. *bootstrap*) procedūra. Ši procedūra naudoja Burrowso deltos plėtinis (Argamon, 2008; Eder *et al.*, 2014) kartu su kilpiniais sutarimo medžiais (angl. *bootstrap consensus trees*). Hierarchinei klasterinei analizei stinga standartizuotų rezultatų patikros metodų (paprastai apsiribojama gautos dendrogramos analize plika akimi), o kilpos procedūra, sujungianti hierarchinės klasterinės analizės dendrogramas ir sprendimų medžius, yra naudingas metodas vertinant rezultatus.

EKSPERIMENTŲ REZULTATAI

Tyrimo rezultatai parodė, jog kalbinės raiškos, žyminčios politinę orientaciją, apimtis labai skiriasi ir daro poveikį skirtingam parlamentinių debatų metu vartojamam žodynui. Tam tikrų skirtumų galima pastebėti jau lyginant vidutinio sakinio ilgį (1 lentelė) ir leksinę įvairovę (2 lentelė). Politinės dešinės vidutinis sakinio ilgis aiškiai ilgesnis (14,42 žodžio) nei politinio centro ir politinės kairės (12,54 žodžio ir 12,81 žodžio atitinkamai). Kita vertus, politinis centras pasižymi didesne leksine įvairove nei politinė kairė ir dešinė, kurių leksinė įvairovė yra gana panaši.

1 lentelė. Vidutinis sakinio ilgis pagal politinę orientaciją

Politinė orientacija	Vidutinis sakinio ilgis
Dešinė	14,42
Centras	12,54
Kairė	12,82

Dažniausių žodžių kaip skiriamųjų požymių rezultatai parodė, jog parlamentarijai, priklausantys politinėms partijoms, priskirtoms skirtingai politinei orientacijai, kalba skirtingai (1 paveikslas). 1 paveiksle vaizduojami stilometrinės analizės rezultatai, kur naudota 10 000 dažniausių žodžių (skiriamieji požymiai), minėtoji Ederio delta ir hierarchinis klasterizavimas. Aiškiai matoma skirtis tarp tekstų pavyzdžių,

² Nėra nusistovėjusio vertimo šiam terminui (angl. *bootstrap*), vertimas *pradinė įkeltis* labiau tinka operacinių sistemų kontekste, šiuo atveju *kilpa* ganėtinai gerai atitinka proceso idėją.

priklausančių skirtingai politinei orientacijai. Be to, politinės dešinės ir politinės kairės tekstai yra panašesni tarpusavyje (sujungti į bendrą klasterį) nei politinio centro tekstai (atskiras klasteris). Siekiant patikrinti rezultatų patikimumą, pasi-
telkta kilpos procedūra. Jai naudota nuo 100 iki 10 000 dažniausių žodžių (atlikta daug iteracijų naudojant skirtingą skiriamųjų požymių – dažniausių žodžių – skai-
čių ir kiekvienai iteracijai sugeneruojant virtualią hierarchinės klasterinės analizės dendrogramą).

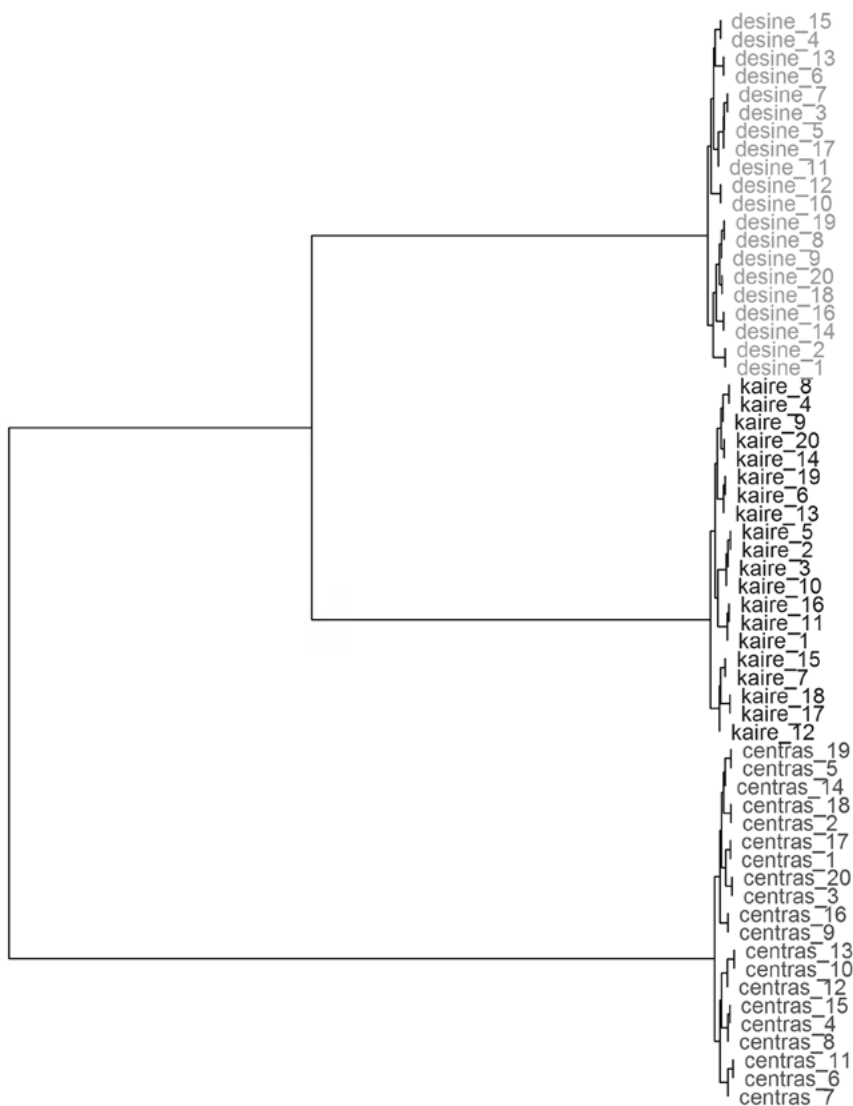
2 lentelė. Leksinė įvairovė pagal politinę orientaciją

Politinė orientacija	Leksinė įvairovė (<i>Standardized Type Token Ratio</i>)
Dešinė	0,0327
Centras	0,0436
Kairė	0,0362

2 paveiksle pateikti rezultatai, kurie parodo ryšius (panašumus) tarp tekstų pavyzdžių, egzistuojančių bent 50 proc. minėtų virtualių dendrogramų. Taigi galima tarti, jog 1 paveiksle atvaizduoti tekstinių pavyzdžių panašumai tikrai egzis-
tuoja.

Norint pažvelgti į konkretesnius politinės kairės, centro ir dešinės kalbinės raiš-
kos ypatumus, pateikiama 20 dažniausių žodžių ir jų dažnių vidurkiai pagal poli-
tinę orientaciją, normalizuoti Z-įverčiais (3 lentelė).

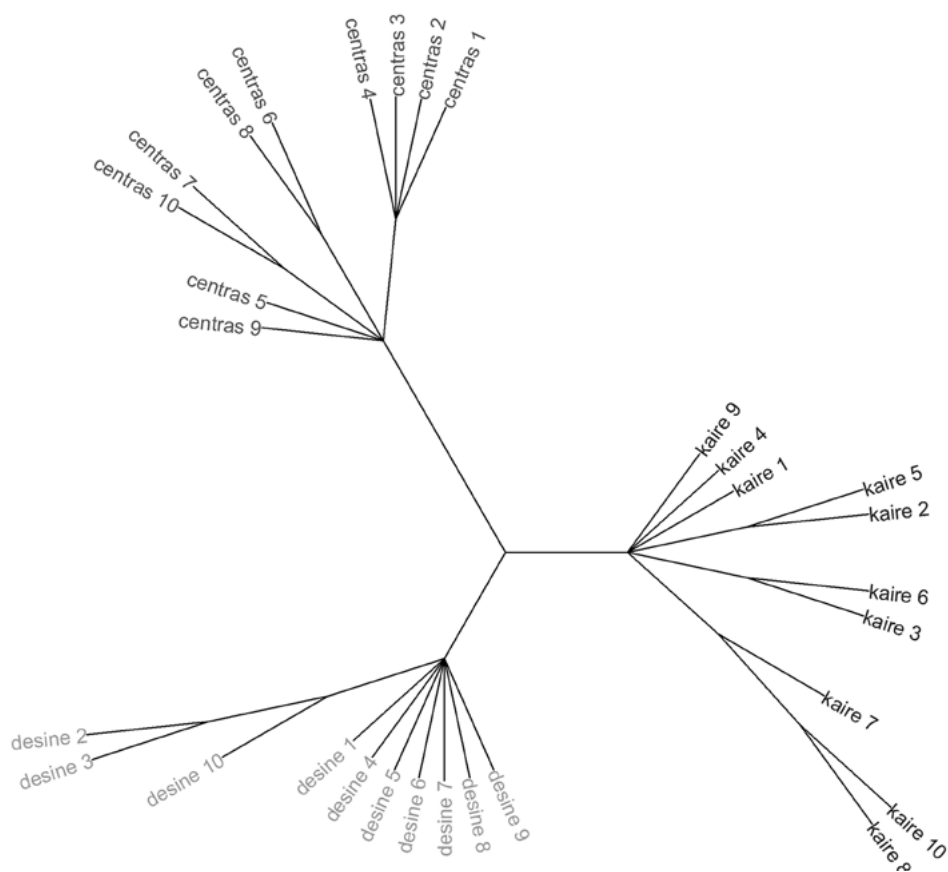
Nors didžiumą lentelės sudaro tarnybiniai / funkciniai žodžiai, tačiau esama ir kitų politikos sričiai charakteringų žodžių. Kalbos technologijų (angl. *Natural Language Processing*) kontekste šie itin dažni žodžiai – tarnybiniai / funkciniai, tam tikram domenui itin charakteringi – laikomi nereikšminiais žodžiais (angl. *stopwords*) ir dau-
geliu atveju iš analizuojamų tekstų yra pašalinami, tačiau atliekant stilometrinę ana-
lizę jie leidžia pamatyti kalbėjimo skirtumus atsižvelgiant į politinės orientacijos „sti-
lių“. Pavyzdžiui, politinė kairė ir dešinė dažniau, palyginti su politiniu centru, vartoja
žodį *mes*, žodį *įstatymo* dažniausiai vartoja politinis centras, rečiausiai – politinė kairė.
Skirtumų esama tarp kitų sąraše esančių žodžių vartojimo – politinė dešinė dažniau
vartoja žodį *ir*, politinė kairė – *kad*, politinis centras – *dėl*. Kadangi į kalbinės raiškos
skirtumus atsižvelgiant į politinę orientaciją žvelgiama per stiliaus prizmę, skirtumai
tarp minėtų nereikšminių žodžių vartojimo yra reikšmingi, nes dažnai stilometri-
nei analizei kaip skiriamieji požymiai naudojami būtent šie žodžiai, siekiant išvengti
temos ir žanro įtakos (Stamatatos, 2009).



1 pav. Hierarchinis klasterizavimas naudojant 10 000 dažniausių žodžių

Siekiant 1 paveiksle vaizduojamų rezultatų detalumo, skirtys tarp politinės dešinės, politinės kairės ir politinio centro rastos atliekant toliau aprašytus veiksmus. Pirma, stenogramų tekstyno pavyzdžiai (1 pavyzdį sudaro 1 Seimo nario pasisakymas nuo jo kalbėjimo pradžios iki pertraukimo arba pabaigos, kada pradeda kalbėti kitas Seimo narys), priklausantys skirtingai politinei orientacijai, sujungiami į

didelius dokumentus atsižvelgiant į jų priskyrimą politinei kairei, politiniam centrui arba politinei dešinei, paskui padalijami į vienodo dydžio mažesnius dokumentus, kad tekstų apdorojimas kompiuteriu vyktų sparčiau. Antra, naudojant visus analizuojamų tekstų pavyzdžius, suformuotas dažninis analizuojamų tekstų žodžių sąrašas, vėliau normalizuotas Z-įverčiais (plačiau žr. „Tyrimo metodas“). Sąrašas sudarytas iš 10 000 dažniausių žodžių per visus tekstų pavyzdžius (dalis jų pateikta 3 lentelėje). Nepaisant daugelio tyrimų, atliktų pasitelkus stilometriją, kol kas nėra universalus sprendimo pasirenkant skiriamųjų požymių kiekį, o 10 000 dažniausių požymių (šiuo atveju žodžių) – vienas galimų pasirinkimų (Stamatatos, 2009). Taikant Ederio deltos atstumo matą (plačiau žr. „Tyrimo metodas“) ir atsižvelgiant į normalizuotus žodžių dažnius randamos tekstų porų nepanašumo vertės. Tuomet sugeneruojama atstumų / distancijų matrica su visomis nepanašumo vertėmis tarp tekstų porų. Galiausiai tekstinių pavyzdžių atvaizdavimui jų panašumo–nepanašumo vartojamų žodžių atžvilgiu taikomas hierarchinis klasterizavimas, kurio dendrograma matoma 1 paveiksle.



2 pav. Kilpinio sutarimų medžio dendrograma su 50 proc. viršijančiais panašumais

2 paveiksle vaizduojami rezultatai, rodantys ryšius (panašumą–nepanašumą) tarp tekstų pavyzdžių pagal politinę orientaciją, atlikus kilpos (angl. *bootstrap*) procedūrą. Į 3 didelius dokumentus atsižvelgiant į politinę orientaciją sujungti tekstų pavyzdžiai buvo padalyti į 10 mažesnių dokumentų, taip analizei turint po 10 dokumentų kiekvienai politinei orientacijai – politinei kairei, politiniam centrui ir politinei dešinei. Naudojant nuo 100 iki 10000 dažniausių žodžių (dalis jų pateikta 3 lentelėje), jų dažnius normalizavus Z-įverčiais (plačiau žr. „Tyrimo metodas“), atliekama daug atstumų tarp tekstų porų matavimo iteracijų (pakartojimų) su skirtingu kiekiu skiriamųjų požymių, kiekvienos jų metu sugeneruojant po virtualią hierarchinės klasterinės analizės dendrogramą. Tuomet pasitelktas sprendimų medžių algoritmas, leidęs atrinkti ir kilpiniais sutarimo medžiais (angl. *bootstrap consensus trees*) atvaizduoti panašumo–nepanašumo ryšius pagal žodžių vartojimą, esančius bent 50 proc. minėtų virtualių dendrogramų. Arčiau ir (ar) ant artimesnių šakų išdėstyti analizuotų tekstų pavyzdžiai yra panašesni nei esantys toliau ir (ar) ant tolimesnių šakų.

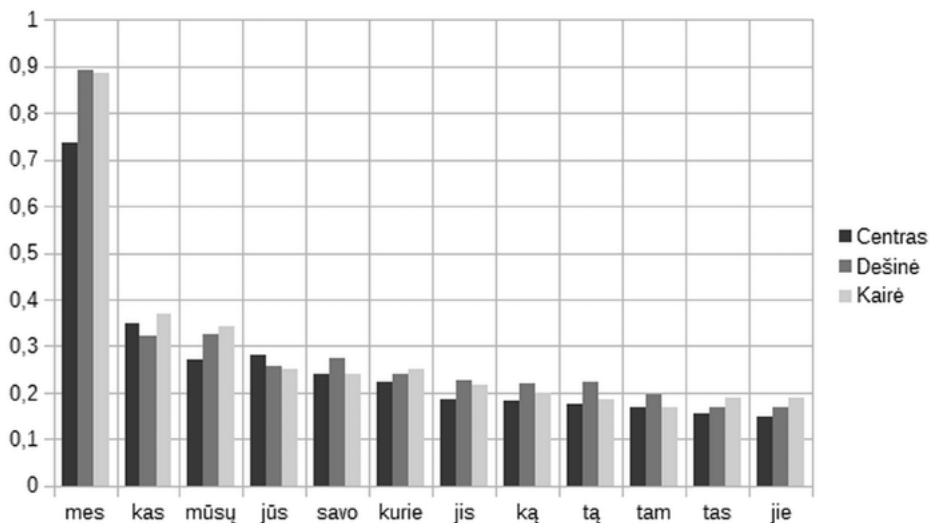
3 lentelė. Dažniausi žodžiai su Z-įverčiais normalizuotais dažniais

	Kairė	Centras	Dešinė
ir	3,36	3,4	3,68
kad	1,92	1,82	1,87
yra	1,45	1,54	1,52
tai	1,20	1,11	1,25
mes	0,89	0,74	0,89
dėl	0,80	0,91	0,73
ar	0,71	0,84	0,68
į	0,65	0,7	0,66
seimo	0,65	0,64	0,54
kaip	0,61	0,57	0,64
įstatymo	0,46	0,71	0,53
už	0,53	0,63	0,44
taip	0,49	0,49	0,50
buvo	0,51	0,45	0,50
o	0,54	0,40	0,49
bet	0,49	0,46	0,48
jeigu	0,48	0,39	0,47
su	0,43	0,44	0,45
ne	0,41	0,42	0,41
būtų	0,43	0,39	0,41

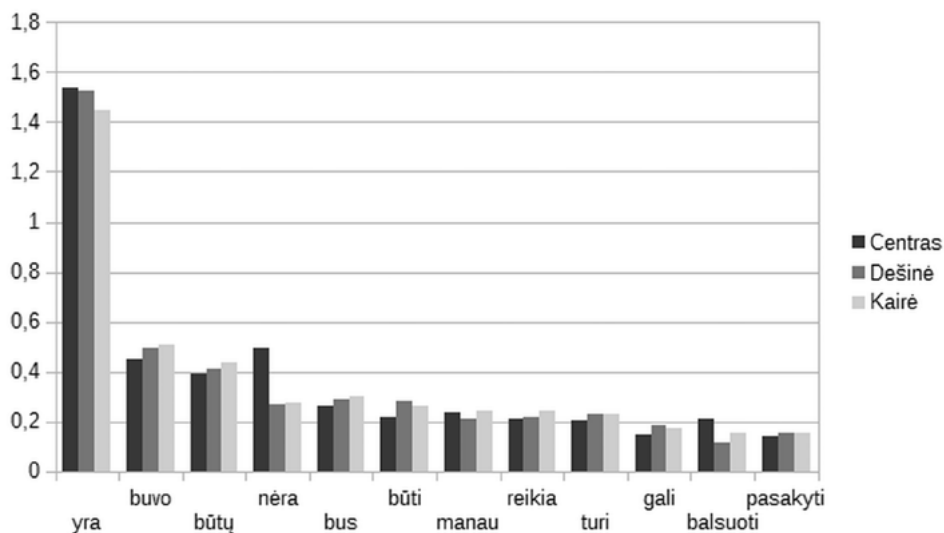
Tęsiant tyrimą ir norint rasti detalesnius skirtumus tarp politinės kairės, centro ir dešinės vartojamo žodyno, išskirti dažniausi įvardžiai (3 paveikslas), veiksmažodžiai (4 paveikslas) ir daiktavardžiai (5 paveikslas). 3, 4 ir 5 paveiksluose atvaizduoti

žodžių (įvardžių, veiksmažodžių ir daiktavardžių) ir jų dažnių vidurkiai pagal politinę orientaciją, normalizuoti Z-įverčiais. Kaip jau buvo minėta, politinis centras rečiau, palyginti su politine kaire ir dešine, vartoja *mes*, tačiau dažniau – *jūs*. Politinės dešinės žodyne dažniau randami *savo*, *jis*, *ką*, *tam*, *tą*, o politinė kairės leksikoje dažniau pasitaiko *kas*, *mūsų*, *tas*, *jie*. Įvardžiai svarbūs tuo, jog, būdami funkcinių / tarnybinių žodžių dalimi (Bell *et al.*, 2009) gali parodyti kalbėtojo santykį su auditorija, su savimi ir tema, kuria kalbama (Pennebaker, 2011).

Tarp dažniausių veiksmažodžių didžiąją dalį sudaro įvairios veiksmažodžio *būti* formos. Ir čia matomi jų vartojimo skirtumai, atsižvelgiant į politinę orientaciją. Politinė kairė rečiau vartoja *yra*, politinis centras išsiskiria veiksmažodžių *nėra* ir *balsuoti* vartojimu, o politinė dešinė – veiksmažodžiais *būti* ir *gali*. Politinės kairės žodyne dažniau aptinkamas *reikia*. Politinės dešinės leksikoje retesnis veiksmažodis *manau*. Veiksmažodžiai perteikia veiksmą, įvykį / atsitikimą (angl. *occurrence*) arba būklę / būseną (Jackendoff, 2002) ir tuo yra svarbūs analizuojant kalbinę raišką. Kita vertus, Seimo narių kalbėjimas gali priklausyti ir nuo jų buvimo parlamentinėje pozicijoje arba parlamentinėje opozicijoje, kai jie turi skirtingą įtaką Seime vykstantiems procesams ir todėl gali naudoti skirtingas kalbinės raiškos priemones, norėdami pasiekti savo tikslų, pvz., įtikinti Seimo narių daugumą balsuoti už tam tikros įstatymo pataisos priėmimą. Šią prielaidą planuojama patikrinti ateityje.

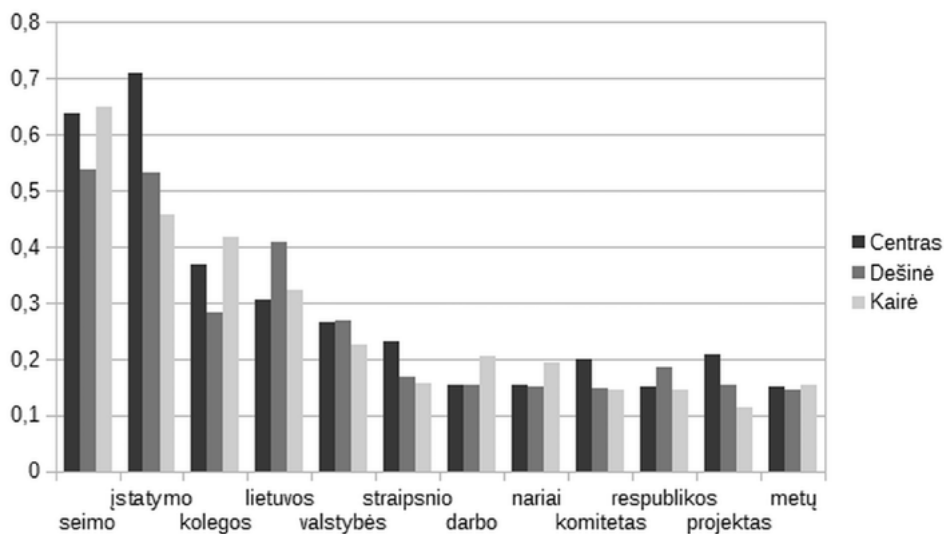


3 pav. Dažniausi įvardžiai, normalizuoti Z-įverčiais



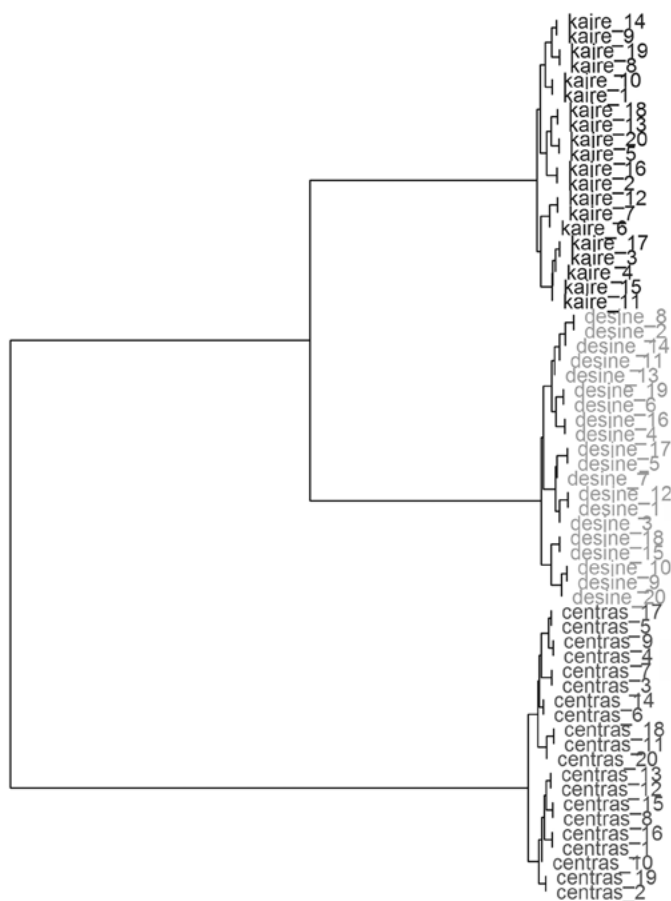
4 pav. Dažniausi veiksmažodžiai, normalizuoti Z-įverčiais

Kalbant apie daiktavardžių vartojimą, politinis centras išsiskiria daiktavardžių *įstatymo* ir *projekto* vartojimu. Politinė dešinė, palyginti su politiniu centru ir politine kaire, išsiskiria retesniu *seimo*, bet dažnesniu *Lietuvos*, *respublikos* vartojimu, o politinė kairė – dažnesniu daiktavardžių *kolegos*, *darbo*, *nariai*. Politinio centro požiūrį į tam tikrus klausimus šiuo atveju sunkiau aptikti, tačiau politinės kairės (*kolegos*, *darbo*, *nariai*) ir dešinės (*Lietuvos*, *respublikos*) nuostatos matomos gana aiškiai.



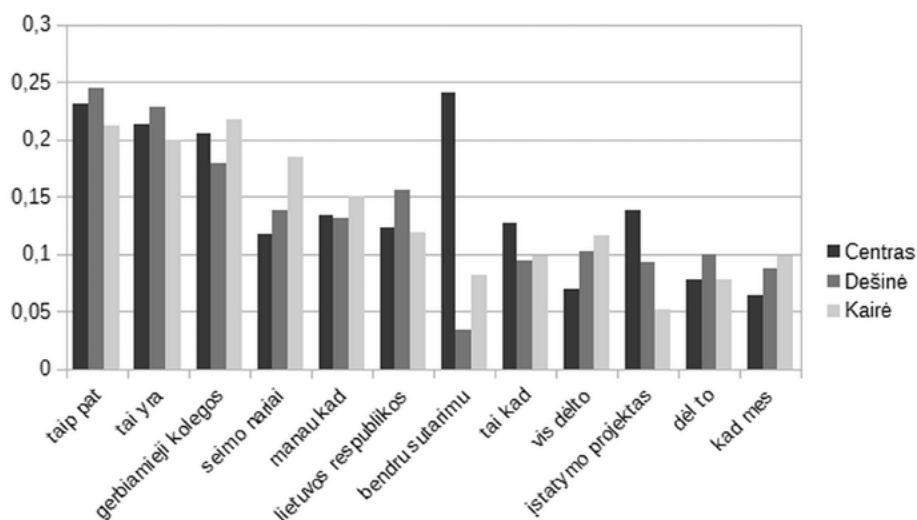
5 pav. Dažniausi daiktavardžiai, normalizuoti Z-įverčiais

Siekiant geresnės rezultatų interpretacijos ir kartu laikantis prielaidos, jog parlamentarų kalba yra gana formalizuota, pasižymi tam tikrų žodžių junginių vartojimu, tyrimui naudotos ir žodžių bigramos. Rezultatai, gauti naudojant žodžių bigramas kaip skiriamuosius požymius, sutampa su Davido Hooverio gautais rezultatais (Hoover, 2002; Hoover, 2003), kadangi dažniausios žodžių bigramos kartu su distancijos matu ir hierarchiniu klasterizavimu sėkmingai leido aptikti kalbinės raiškos panašumus ir skirtumus atsižvelgiant į parlamentarų partijų, kurioms jie priklauso, politinę orientaciją. Kita vertus, Ederio (2011) tyrimų rezultatai naudojant žodžių bigramas ir trigramas buvo sėkmingi tiriant angliškų tekstus, tačiau kitų kalbų, ypač lotynų, tekstams – beveik nie. Rezultatai naudojant žodžių bigramas pateikti 6 paveiksle. Jame vėlgi aiškiai matyti kalbinės raiškos skirtumai, o politinės kairės ir politinės dešinės leksika yra panašesnė nei politinio centro.



6 pav. Hierarchinis klasterizavimas naudojant 5000 dažniausių žodžių bigramų

7 paveiksle matyti, kad politinis centras itin pasižymi junginio *bendru sutarimu* vartojimu. *Išstatymo projektas* taip pat gana dažnas šiai politinei orientacijai priskirtuose analizuojamuose teksto pavyzdžiuose. Politinės dešinės žodyne dažnesni junginiai *taip pat, tai yra, Lietuvos Respublikos, dėl to*. Tuo tarpu politinės kairės leksikoje randami junginiai: *gerbiamieji kolegos, Seimo nariai, manau kad, vis dėlto*. Nemaža dalis dažniausių žodžių bigramų yra morfologinės samplaikos (Rimkutė ir Kovalevskaitė, 2010), kurias būtų galima priskirti prie tarnybinių / funkcinių junginių, kaip ir tarnybiniai žodžiai parodančių kalbėtojo santykį su auditorija, tema, kuria kalbama. Kitos frazės yra standartiniai procedūriniai žodžių junginiai, įprasti Seimo darbo kontekste. Vis dėlto jie taip pat parodo tam tikrą požiūrį į darbą Seime, atliekamą vaidmenį etc., kad ir junginiai *gerbiamieji kolegos* ar *bendru sutarimu*.



7 pav. Dažniausios žodžių bigramos, normalizuotos Z-įverčiais

IŠVADOS

Stilometrinė analizė, nors ir nėra įprasta kalbinės raiškos tyrimams lietuviškuose tekstuose, leido giliau pažvelgti į Lietuvos Respublikos Seimo narių kalbinės raiškos panašumus ir skirtumus, rasti charakteringesnius žodžių, jų junginių vartojimo pavyzdžius atsižvelgiant į politinę orientaciją (kairė, centras, dešinė). Eksperimentai parodė, kad politinei dešinei priklausantys parlamentarai kalba vidutiniškai

ilgesniais sakiniais nei politinis centras ar politinė kairė. Kita vertus, politinis centras pasižymi didesne leksine įvairove, o politinės dešinės ir politinės kairės leksika yra panaši.

Net ir nedidelė dalis dažniausių žodžių ar jų junginių, priskirtinų tarnybiniams / funkciniais žodžiams (ar jų junginiams), leido nustatyti politinės kairės, centro ir dešinės kalbėjimo skirtumus. Tarnybiniai / funkciniai žodžiai, skirtingai nuo turinio žodžių, savyje turi nedaug leksinės reikšmės (arba ji yra nevienareikšmė), tačiau nusako ryšius tarp žodžių sakinyje arba parodo kalbėtojo požiūrį. Pavyzdžiui, politinės kairės ir politinės dešinės žodyne dažnesnis *mes*. Be to, minėtų tarnybinių / funkcinų žodžių ar junginių vartoseną pasirodė reikšminga tiriant kalbinę raišką, reiškiant įsitraukimą, tam tikrą bendrumą, o politinio centro pavyzdžiuose dažnesnis *jūs* – perteikia atsiribojimą, atskirumą.

Kita vertus, Seimo narių kalbėjimas, jo ypatumai gali priklausyti ne tik nuo ideologijos ir politinio spektro. Parlamentarų kalbėjimui, raiškos priemonių pasirinkimui įtakos gali turėti jų priklausymas parlamentinei pozicijai arba opozicijai, galios ir įtakos Seime dinamika. Apie tai galėtų signalizuoti, pvz., veiksmažodžių ir jų formų vartojimas (politinės kairės *reikia*, politinės dešinės *gali* ir politinio centro *nera*). Parlamentinės pozicijos ir opozicijos kalbinę raišką planuojama tirti ateityje. Be to, atlikti stilometrinę analizę planuojama atsižvelgiant ir į kitus pjūvius – atskirų Seimo narių kalbėjimą, kalbėjimo kitimą per laiką, atsižvelgti į lytį, išsilavinimą ir kt.

Dažniausios žodžių bigramos leido užfiksuoti tam tikrą stereotipinę kalbą (angl. *formulaic language*) – būdingus žodžių junginius, nusakančius kalbėtojo santykį su auditorija, savimi ir tema (morfologinės samplaikos), mandagumo frazes, kitus procedūrinius žodžių junginius. Politinės kairės leksikoje dažniau pasirodantys *gerbiamieji kolegos*, *Seimo nariai*, politinio centro – *bendru sutarimu* yra gana sustabarėję junginiai, dažni Seimo posėdžių kalboje. Politinės dešinės kalboje dažnesnės buvo morfologinės samplaikos *taip pat* (tęstinumas), *tai yra* (patikslinimas), *dėl to* (paaiškinimas, priežastingumas), o politinės kairės – junginys *manau kad* nuomonės išsakymui (skyrybos ženklai buvo pašalinti pirminiame tekstų apdorojimo etape) bei morfologinė samplaika *vis dėlto* (priešpriešinimas).

Taigi stilometrinių analizė sėkmingai pasitarnavo Lietuvos Respublikos Seimo posėdžių stenogramų analizei, siekiant ištirti parlamentarų kalbinę raišką žodžių ir jų junginių vartojimo atžvilgiu. Tačiau toliau interpretuojant rezultatus reikia atsižvelgti į tai, kad duomenys paimti iš labai ilgo laiko intervalo, per kurį kalbėsena galėjo keistis, todėl būtų labai įdomu pažiūrėti, kaip stilius keitėsi, t. y. ar partijų (frakcijų) retorika nuo 1990 m. pasikeitė. Kitas svarbus pjūvis, kuris nebuvo analizuotas šiame tyrime, yra buvimas pozicijoje arba opozicijoje – šiuo

metu duomenyse tai neišskirta, nors tikėtina, kad opozicijos pasisakymai gali būti labiau negatyvūs, o pozicijos – pozityvūs. Kiti įdomūs pјjviai yra parlamentarų amžius, lytis ir išsilavinimas. Tolesni tyrimų planai įtraukia visų šių faktorių įtaką kalbinei raiškai.

LITERATŪRA

- Antonia, Alexis, Craig, Hugh, and Elliott Jack. Language chunking, data sparseness, and the value of a long marker list: explorations with word n-grams and authorial attribution. *Literary and Linguistic Computing* 29(2), 2014, 147–163.
- Argamon, Shlomo, Koppel, Moshe, Fine, Jonathan, and Shimoni, Anat Rachel. Gender, genre, and writing style in formal written texts. *Text*, 23(3), 2003, 321–346.
- Argamon, Shlomo, Whitelaw, Casey, Chase, Paul, Hota, Sobhan Raj, Garg, Navendu, and Levitan Shlomo. Stylistic text classification using functional lexical features. *Journal of the American Society for Information Science and Technology*, 58(6), 2007, 802–822.
- Argamon, Shlomo. Interpreting Burrows’s Delta: Geometric and probabilistic foundations. *Literary and Linguistic Computing*, 23(2), 2008, 131–147.
- Bell, Alan, Brenier, Jason M., Gregory, Michelle, Girand, Cynthia and Jurafsky, Dan. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 2009, 92–111.
- Bykh, Serhiyand and Meurers, Detmar. Native language identification using recurring n-grams – investigating abstraction and domain dependence. *COLING*, 2012, 425–440.
- Brennan, Patrick. Uses of computational stylometry to determine demographics for online reputation management. *Language Engineering for Online Reputation Management*, 2012, 15.
- Burrows, John F. ‘Delta’: A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 17(3), 2002, 267–287.
- Burrows, John F. Not unless you ask nicely: The interpretative nexus between analysis and information. *Literary and Linguistic Computing*, 7(2), 1992, 91–109.
- Celli, Fabio Unsupervised personality recognition for social network sites. In *ICDS 2012, The Sixth International Conference on Digital Society*, 2012, 59–62.
- Cheng, Na, Chen, Xiaoling, Chandramouli, Rajarathnam, and Subbalakshmi, K. P. Gender identification from e-mails. *IEEE Symposium on Computational Intelligence and Data Mining, CIDM’09*, 2009, 154–158.
- Daelemans, Walter. Explanation in computational stylometry. *Computational Linguistics and Intelligent Text Processing*, Springer, 2013, 451–462.
- Dahllöf, Mats. Automatic prediction of gender, political affiliation, and age in Swedish politicians from the wording of their speeches – a comparative study of classifiability. *Literary and linguistic computing*, 27(2), 2012, 139–153.
- De-Arteaga, Maria, Jimenez, Sergio, Duenas, George, Mancera, Sergio, and Baquero, Julia. Author profiling using corpus statistics, lexicons and stylistic features. *Online Working Notes of the 10th PAN Evaluation Lab on Uncovering Plagiarism, Authorship and Social Misuse, CLEF*, 2013.
- Eder, Maciej, Rybicki, Jan, and Kestemont, Mike. Package ‘stylo’. 2014.
- Eder, Maciej and Rybicki, Jan. Do birds of a feather really flock together, or how to choose training samples for authorship attribution. *Literary and Linguistic Computing*, 28(2), 2013, 229–236.
- Eder, Maciej Does size matter? Authorship attribution, small samples, big problem. *Proceedings of Digital Humanities*, 2010, 132–135.
- Eder, Maciej. Computational stylistics and biblical translation: How reliable can a dendrogram be. *The translator and the computer*, 2013a, 155–170.
- Eder, Maciej. Mind your corpus: systematic errors in authorship attribution. *Literary and linguistic computing*, 28(4), 2013b, 603–614.
- Eder, Maciej. Style-markers in authorship attribution: a cross-language study of authorial fingerprint. *Studies in Polish Linguistics*, 2011, 6, 99–114.
- Gaivenis, Kazimieras, Keinys, Stasys. *Kalbotoyros terminų žodynas*. Kaunas, Šviesa, 1990.
- Galinskaitė, Reda. Pritarimo strategija jaunimo pokalbiuose: funkcijos ir raiška. *Lietuvių kalba*, 5, 2014.
- Garrard, Peter. Cognitive archaeology: Uses, methods, and results. *Journal of Neurolinguistics*, 22(3), 2009, 250–265.

- Gianfortoni, Philip, Adamson, David, and Rosé, Carolyn P. Modeling of stylistic variation in social media with stretchy patterns. *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties, Association for Computational Linguistics*, 2011, 49–59.
- Girčienė, Jurgita. Daugiskaitinių ir vienaskeitinių kreipimosi formų konkurencija viešojoje sakytinėje vartosenoje (sociolingvistinis aspektas). *Kalbos kultūra*, 82, 2009, 214–234.
- Goren, Paul. Party identification and core political values. *American Journal of Political Science*, 49(4), 2005, 881–896.
- Green, Donald, Palmquist, Bradley, and Schickler, Eric. *Partisan hearts and minds*. Yale University Press, 2002, 720–744.
- Halteren, Hans Van, Baayen, Harald, Tweedie, Fiona, Haverkort, Marco, and Neijt, Anneke. New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics*, 12(1), 2005, 65–77.
- Hochmann, Jean-Rémy, Endress, Ansgar D., and Mehler, Jacques. Word frequency as a cue for identifying function words in infancy. *Cognition*, 115(3), 2010, 444–457.
- Holmes, David I., Gordon, Lesley J., and Wilson, Christine. A widow and her soldier: Stylometry and the american civil war. *Literary and Linguistic Computing*, 16(4), 2001, 403–420.
- Hoover, David L. Corpus stylistics, stylometry, and the styles of Henry James. *Style*, 41(2), 2007, 174.
- Hoover, David L. Delta prime? *Literary and Linguistic Computing*, 19(4), 2004, 477–495.
- Hoover, David L. Frequent collocations and authorial style. *Literary and Linguistic Computing*, 18(3), 2003, 261–286.
- Hoover, David L. Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing*, 17(2), 2002, 157–180.
- Jackendoff, Ray. *Foundations of Language*, Oxford University Press, 2002.
- Juola, Patrick, and Baayen, R. Harald. A controlled-corpus experiment in authorship identification by cross-entropy. *Literary and Linguistic Computing*, 20 (Suppl), 2005, 59–67.
- Kapočiūtė-Dzikiėnė, Jurgita and Krupavičius, Algis. Predicting party group from the Lithuanian parliamentary speeches. *Information Technology and Control*, 43(3), 2014, 321–332.
- Kapočiūtė-Dzikiėnė, Jurgita, Šarkutė, Ligita ir Utkā, Andrius. Seimo posėdžių stenogramų tekstynas autorystės nustatymo bei autoriaus profilio sudarymo tyrimams. *Linguistics: Germanic & Romance Studies / Kalbotyra: Romanų ir Germanų Studijos*, 66, 2014a.
- Kapočiūtė-Dzikiėnė, Jurgita, Šarkutė, Ligita, and Utkā, Andrius. Automatic Author Profiling of Lithuanian Parliamentary Speeches: Exploring the Influence of Features and Dataset Sizes. *Human Language Technologies – The Baltic Perspective: proceedings*, IOS press, 2014c, 99–106.
- Kapočiūtė-Dzikiėnė, Jurgita, Utkā, Andrius, and Šarkutė, Ligita. Feature exploration for authorship attribution of Lithuanian parliamentary speeches. *In Text, Speech and Dialogue*, Springer, 2014b, 93–100.
- Kestemont, Mike. Function words in authorship attribution from black magic to theory? *EACL*, 2014, 59–66.
- Kinder, Donald R. *Opinion and action in the realm of politics*. McGraw-Hill, 1998.
- Koppel, Moshe, Argamon, Shlomo, and Shmoini, Anat Rachel. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17(4), 2002, 401–412.
- Luyckx, Kim and Daelemans, Walter. Personae: a corpus for author and personality prediction from text. *In LREC*, 2008.
- Mandravickaitė, Justina. Stylometry for visualizing rhetorical differences in parliamentary speech. *In XVIII tarptautinės tarptautinės magistrantų ir doktorantų konferencijos „Informacinė visuomenė ir universitetinės studijos“ (IVUS 2013) medžiaga*, 2013, 68–72.
- Mechti, Seifeddine, Jaoua, Maher, and Belguith, Lamia Hadrach. Author profiling using style-based features. *Notebook for PAN at CLEF*, 2013.
- Morgan, G. Scott, Mullen, Elizabeth, and Skitka, Linda J. When values and attributions collide: Liberals' and conservatives' values motivate attributions for alleged misdeeds. *Personality and Social Psychology Bulletin*, 2010.
- Mosteller, Frederick and Wallace, David. *Inference and disputed authorship: The Federalist*. Addison-Wesley, 1964.
- Nguyen, Dong, Smith, Noah A., and Rosé Carolyn P. Author age prediction from text using linear regression. *In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, Association for Computational Linguistics*, 2011, 115–123.
- Pennebaker, James W. The secret life of pronouns. *New Scientist*, 211 (2828), 2011, 42–45.
- Pikčilingis, Juozas. *Kas yra stilius*. Vilnius, Vaga, 1971.
- Rashid, Awas, Rayson, Paul, Greenwood, Phil, Walkerdine, James, Duquenoy, Penny, Watson, Patrick, Brennan, Margaret, and Jones, Matt. Isis: protecting children in online social networks. 2009.
- Rybicki, Jan and Eder, Maciej. Deeper Delta across genres and languages: do we really need the most frequent words? *Literary and linguistic computing*, 26(3), 2011, 315–321.

- Rimkutė, Erika ir Kovalevskaitė, Jolanta. Sudėtinės ir suaugtinės lietuvių kalbos morfologinės samplaiškos. *Kalbų studijos*, 16, 2010, 79–88.
- Rustagi, Mayur, Prasath, R. Rajendra, Goswami, Sumit, and Sarkar, Sudeshna. Learning age and gender of blogger from stylistic variation. In *Pattern Recognition and Machine Intelligence*, Springer, 2009, 205–212.
- Sarawgi, Ruchita, Gajulapalli, Kailash, and Choi, Yejin. Gender attribution: tracing stylometric evidence beyond topic and genre. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning, Association for Computational Linguistics*, 2011, 78–86.
- Savickienė, Ineta. Komunikacinė pragmatika ir kalbėjimo situacijos tikslas: deminutyvų vartojimo atvejis. *Kalbos kultūra*, 79, 2006, 256–263.
- Sigurd, Bengt, Eeg-Olofsson, Mats, and Van Weijer, Joost. Word length, sentence length and frequency--Zipf revisited. *Studia Linguistica*, 58 (1), 2004, 37–52.
- Stamatatos, Efstathios. A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, 60(3), 2009, 538–556.
- Šarkutė, Ligita and Utka, Andrius. The effect of author set size in authorship attribution for Lithuanian. In *Nordic Conference of Computational Linguistics NODALIDA*, 2015, 87.
- Žalkauskaitė, Gintarė. Idiolektų požymiai elektroninių laiškų skyryboje. *Lietuvių kalba*, 5, 2014.
- Žalkauskienė, Anelė. Lietuviško teksto autoriaus nustatymo metodikos pagrindai. *Jurisprudencija*, 18, 2000.

Justina Mandravickaitė, Tomas Krilavičius

LANGUAGE USAGE OF MEMBERS OF THE LITHUANIAN PARLIAMENT CONSIDERING THEIR POLITICAL ORIENTATION

SUMMARY. This study presents a stylometric analysis of a corpus of transcribed Lithuanian Parliamentary speeches for the purpose of exploring them according to their political orientation (left-center-right). The methods of computer-aided stylometry applied in the research allow us to describe and explain the connection between the author's psychological and sociological attributes and to analyze his/her style. In this research, political orientation is treated as an attribute/feature exhibited by the members of Parliament under analysis; i.e., the perspective is stylistic and distinguishes between styles of the political left, political center, and political right. The results of the research show that the verbal expressions peculiar to a political orientation range very widely and make a difference to the vocabulary used in Parliamentary debates.

KEYWORDS: transcripts of plenary sittings of the Lithuanian Parliament, stylometry, computational stylistics, political orientation.