

<https://doi.org/10.15388/vu.thesis.72>
<https://orcid.org/0000-0001-9426-6165>

VILNIAUS UNIVERSITETAS
LIETUVIŲ LITERATŪROS IR TAUTOSAKOS INSTITUTAS

Justina
MANDRAVICKAITĖ

Kalbinė funkcinių stilių raiška: kiekybinė požymių analizė

DAKTARO DISERTACIJA

Humanitariniai mokslai,
filologija (H 004)

VILNIUS 2020

Disertacija rengta 2015–2019 metais Vilniaus universitete.

Mokslinė vadovė:

doc. dr. Loreta Vilkienė [Vilniaus universitetas, humanitariniai mokslai, filologija – H 004].

Mokslinis konsultantas:

prof. dr. Tomas Krilavičius [Vytauto Didžiojo universitetas, gamtos mokslai, informatika – N 009].

TURINYS

LENTELIŲ SĄRAŠAS.....	8
PAVEIKSLŲ SĄRAŠAS.....	10
SANTRUMPOS.....	11
SĄVOKŲ ŽODYNĖLIS.....	12
ĮVADAS.....	14
1. TEORINĖS TYRIMO PRIEIGOS.....	27
1.1 Funkcinė stilistika ir funkcinio stiliaus samprata.....	27
1.1.2 FS samprata.....	28
1.1.3 FS samprata Lietuvos stilistikos mokyklos tradicijoje.....	30
1.2 Funkcinių stilių sistema ir jų požymiai.....	31
1.2.1 FS tipologijos.....	31
1.2.2 Lietuvių kalbos administracinio stiliaus ypatybės.....	36
1.2.3 Lietuvių kalbos mokslinio stiliaus ypatybės.....	41
1.2.4 Lietuvių kalbos publicistinio stiliaus ypatybės.....	47
1.3 Kiekybinė lingvistika: trumpa istorija.....	50
1.3.2 Ankstyvasis kiekybinės lingvistikos laikotarpis (iki 1935 m.).....	51
1.3.3 Nuo Zipf'o iki Altmann'o (nuo 1935 m. iki 1970 m.).....	52
1.3.4 KL plėtros (nuo 1970 m.) ir Köhler sinergetinės lingvistikos (nuo 1993 m.) laikotarpiai.....	53
1.4 Kiekybiniai indikatoriai FS kalbinės raiškos tyrime.....	56
1.4.1 Zipf'o dėsnis.....	57
1.4.2 Teksto dažnine struktūra paremti kiekybiniai indikatoriai.....	57
1.4.2.1 Indikatorius a	58
1.4.2.2 Indikatorius R_1	60
1.4.2.3 Indikatorius RR_{mc}	61
1.4.2.4 Indikatorius MATTR.....	63
1.4.2.5 Teminės koncentracijos indikatorius.....	65
1.4.3 Kiti indikatoriai.....	68

1.4.3.1 Vidutinis žodžių ir žodžių formų ilgis (VŽFI)	68
1.4.3.2 Aktyvumo (A) ir deskriptyvumo (D) indikatoriai	69
1.4.2.3 Tarpveiksmažodiniai atstumai (VA)	71
1.5 Tyrimų, kuriuose taikyti kiekybiniai indikatoriai tekstų savybėms tirti, apžvalga.....	73
1.5.1 Metodologinio pobūdžio publikacijų apžvalga.....	73
1.5.2 Trumpa taikomojo pobūdžio literatūros apžvalga.....	76
2. TYRIMO MEDŽIAGA.....	85
2.1 Tiriamosios medžiagos pagrindas	85
2.2 Tiriamosios medžiagos parengimo žingsniai	86
2.3 Tikslinių tekstynų sandara.....	90
2.3.1 Administracinio FS tekstynas	90
2.3.2 Mokslinio FS tekstynas	90
2.3.3 Publicistinio FS tekstynas	92
3. TYRIMO METODAI IR EIGA	93
3.1 Aprašomoji duomenų analizė: metodologinės pastabos	93
3.2 Koreliacinė analizė: metodologinės pastabos	96
3.3 Pagrindinių komponentų analizė: metodologinės pastabos.....	98
3.4 Dispersinė analizė: metodologinės pastabos.....	100
3.5 Tyrimo eiga.....	102
4. ATSKIRŲ FS KALBINĖS RAIŠKOS POŽYMIŲ ANALIZĖ.....	103
4.1 Administracinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė.....	103
4.1.1 Aprašomoji (pirminė) administracinių tekstų (AST) požymių analizė	103
4.1.1.1 Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus savybės AST	104
4.1.1.2 Indikatoriaus a savybės AST.....	105
4.1.1.3 R_1 indikatoriaus savybės AST	107
4.1.1.4 AST tekstų RR_{mc} indikatoriaus charakteristikos	109
4.1.1.5 MATTR indikatoriaus charakteristikos AST	110

4.1.1.6	AST teminės koncentracijos (TK) indikatoriaus charakteristikos.	111
4.1.1.7	AST tekstų indikatorių A ir D charakteristikos	112
4.1.1.8	Tarpveiksmažodinių atstumų (VA) indikatoriaus savybės administraciniuose tekstuose	114
4.1.2	Indikatorių koreliacija administracinio stiliaus tekstams	115
4.1.2.1	Išskirtys administracinio stiliaus tekstyno duomenyse.....	115
4.1.2.2	Koreliacinės analizės rezultatai AST.....	117
4.1.2.2.1	Labai stipri koreliacija tarp indikatorių	119
4.1.2.2.2	Vidutiniškai stipri koreliacija tarp indikatorių.....	119
4.1.2.2.3	Silpna ir labai silpna koreliacija tarp indikatorių.....	121
4.1.3	Pagrindinių komponentų analizės (PKA) rezultatai AST.....	121
4.1.3.1	Indikatorių svarba reprezentuojant administracinio stiliaus kalbinę raišką	121
4.1.3.2	Indikatorių indėlis reprezentuojant administracinio stiliaus kalbinę raišką	125
4.2	Mokslinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė	128
4.2.1	Aprašomoji (pirminė) mokslinių tekstų požymių analizė	128
4.2.1.1	Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus charakteristikos mokslinio stiliaus tekstuose	129
4.2.1.2	Indikatoriaus a charakteristikos MST	130
4.2.1.3	R_1 indikatoriaus charakteristikos MST.....	130
4.2.1.4	MST tekstų RR_{mc} indikatoriaus charakteristikos.....	131
4.2.1.5	MATTR indikatoriaus charakteristikos MST.....	132
4.2.1.6	Teminės koncentracijos (TK) indikatoriaus charakteristikos MST	133
4.2.1.7	MST tekstų indikatorių A ir D charakteristikos	134
4.2.1.8	Tarpveiksmažodinių atstumų (VA) indikatoriaus charakteristikos moksliniuose tekstuose.....	135
4.2.2	Mokslinio stiliaus tekstų indikatorių koreliacija	135
4.2.2.1	Išskirtys MST duomenyse	136
4.2.2.2	Koreliacinės analizės rezultatai MST	137
4.2.2.2.1	Labai stipri koreliacija tarp indikatorių	139

4.2.2.2.2 Vidutiniškai stipri koreliacija tarp indikatorių.....	139
4.2.2.2.3 Silpna ir labai silpna koreliacija tarp indikatorių.....	141
4.2.3 Pagrindinių komponentų analizės (PKA) rezultatai MST	141
4.2.3.1 Indikatorių svarba reprezentuojant mokslinio stiliaus kalbinę raišką	141
4.2.3.2 Indikatorių indėlis reprezentuojant mokslinio stiliaus kalbinę raišką	144
4.3 Publicistinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė	148
4.3.1 Aprašomoji (pirminė) publicistinių tekstų požymių analizė	148
4.3.1.1 Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus charakteristikos PST tekstuose.....	149
4.3.1.2 Indikatoriaus a charakteristikos PST.....	150
4.3.1.3 R_1 indikatoriaus charakteristikos PST	151
4.3.1.4 PST tekstų RR_{mc} indikatoriaus charakteristikos	151
4.3.1.5 MATTR indikatoriaus charakteristikos PST	152
4.3.1.6 Teminės koncentracijos (TK) indikatoriaus charakteristikos PST .	153
4.3.1.7 Indikatorių A ir D charakteristikos PST	154
4.3.1.8 Tarpveiksmažodinių atstumų (VA) indikatoriaus charakteristikos PST tekstuose.....	155
4.3.2 Indikatorių koreliacija PST tekstams.....	156
4.3.2.1 Išskirtys PST duomenyse	157
4.3.2.2 Koreliacinės analizės rezultatai PST	158
4.3.2.2.1 Labai stipri koreliacija tarp indikatorių	160
4.3.2.2.2 Vidutiniškai stipri koreliacija tarp indikatorių.....	161
4.3.2.2.3 Silpna ir labai silpna koreliacija tarp indikatorių.....	162
4.3.3 PST pagrindinių komponentų analizės (PKA) rezultatai.....	163
4.3.3.1 Indikatorių svarba reprezentuojant publicistinio stiliaus kalbinę raišką	163
4.3.3.2 Indikatorių indėlis reprezentuojant publicistinio stiliaus kalbinę raišką	166

4.4	Atskirų FS kalbinės raiškos požymių analizės apibendrinimas	170
4.4.1	Administracinio stiliaus kalbinės raiškos požymių kiekybinė analizės apibendrinimas	172
4.4.2	Mokslinio stiliaus kalbinės raiškos požymių kiekybinė analizės apibendrinimas	175
4.4.3	Publicistinio stiliaus kalbinės raiškos požymių kiekybinė analizės apibendrinimas	179
5.	TIRIAMŪJŲ FS PALYGINIMAS	183
5.1	FS tarpusavio panašumų ir skirtumų pagal kalbinę raišką žyminčius indikatorius analizė.....	183
5.2	Funkcinių stilių tarpusavio panašumų ir skirtumų apibendrinimas .	190
	IŠVADOS.....	193
	LITERATŪROS SĄRAŠAS.....	201
	PRIEDAI	216

LENTELIŲ SĄRAŠAS

1 lentelė. FS tipologijų pavyzdžiai	32
2 lentelė. Indikatoriaus a verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams	59
3 lentelė. Indikatoriaus R_1 verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams.....	60
4 lentelė. Indikatorių R_1 ir RR_{mc} verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams	62
5 lentelė. Indikatorių MATTR ir R_1 verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams.....	64
6 lentelė. Teksto dažninės lentelės fragmentas	66
7 lentelė. Indikatoriaus TK verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams.....	67
8 lentelė. Indikatoriaus VŽFI verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams	69
9 lentelė. Indikatorių A ir D verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams	70
10 lentelė. Indikatoriaus VA verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams	72
11 lentelė. Tikslinių tekstynų tekstynų charakteristikos	89
12 lentelė. Koreliacijos koeficiento reikšmės (pagal Čekanavičių ir Murauską (2001, 126)).....	97
13 lentelė. AST: bazinių statistinių charakteristikų santrauka	103
14 lentelė. AST teksto <i>adm0-100.txt</i> dažninės lentelės fragmentas (pilka spalva žymi h-tašką).....	106
15 lentelė. AST teksto <i>adm81-74.txt</i> dažninės lentelės fragmentas (pilka spalva žymi h-tašką).....	107
16 lentelė. AST: indikatorių verčių išskirtys.....	116
17 lentelė. AST: koreliacijos tarp indikatorių porų statistinis reikšmingumas (<i>p</i> -vertės).....	118
18 lentelė. PKA rezultatų santrauka AST: duomenų variacijos proporcijos, kurias paaiškina komponentės.....	122
19 lentelė. PKA rezultatų santrauka AST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis.....	123
20 lentelė. MST: bazinių statistinių charakteristikų santrauka.....	128
21 lentelė. MST: indikatorių verčių išskirtys	136
22 lentelė. MST: koreliacijos statistinis reikšmingumas (<i>p</i> -vertės)	138
23 lentelė. PKA rezultatų santrauka MST: duomenų variacijos proporcijos, kurias paaiškina komponentės.....	142

24 lentelė. PKA rezultatų santrauka MST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis.....	143
25 lentelė. PST: bazinių statistinių charakteristikų santrauka.....	148
26 lentelė. PST: indikatorių verčių išskirtys	157
27 lentelė. PST: koreliacijos statistinis reikšmingumas (p-vertės).....	160
28 lentelė. PKA rezultatų santrauka PST: duomenų variacijos proporcijos, kurias paaiškina komponentės.....	164
29 lentelė. PKA rezultatų santrauka PST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis.....	165
30 lentelė. Kruskal'o-Wallis'o testo rezultatai.....	183
31 lentelė. Epsilon kvadrato testo rezultatai.....	184
32 lentelė. VDA testo rezultatai	184
33 lentelė. Dunn'o testo rezultatai.....	185
34 lentelė. Santykiniai poveikio įverčiai	186

PAVEIKSLŲ SĄRAŠAS

1 pav. Administracinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu.....	39
2 pav. Mokslinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu	44
3 pav. Publicistinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu.....	48
4 pav. Smuiko diagramos duomenų savybių iliustravimui	94
5 pav. Koreliacinės analizės diagrama AST.....	117
6 pav. AST: indikatorių indėlis (PK1 ir PK2).....	126
7 pav. AST: indikatorių indėlis (PK2 ir PK3).....	127
8 pav. Koreliacinės analizės diagrama MST	137
9 pav. MST: indikatorių indėlis (PK1 ir PK2)	146
10 pav. MST: indikatorių indėlis (PK2 ir PK3)	147
11 pav. Koreliacinės analizės diagrama PST	159
12 pav. Indikatorių indėlis PST: PK1 ir PK2	168
13 pav. PST: indikatorių indėlis: PK2 ir PK3	169
14 pav. Indikatorių verčių pasiskirstymo AST diagrama.....	170
15 pav. Indikatorių verčių pasiskirstymo MST diagrama	171
16 pav. Indikatorių verčių pasiskirstymo PST diagrama	171
17 pav. FS požymius apibendrinanti diagrama	190

SANTRUMPOS

- A – aktyvumo / dinamiškumo įvertis
- AST – administracinio funkcinio stiliaus tekstynas
- D – deskriptyvumo / aprašomumo įvertis
- DLKT – Dabartinės lietuvių kalbos tekstynas
- FS – funkcinis stilius
- KL – kiekybinė lingvistika
- KP – kvartilinis plotis
- MATTR – informacijos srauto įvertis arba vidutinis skirtingų žodžių ir žodžių formų kiekis tekste (angl. Moving Average Type-Token Ratio)
- MST – mokslinio funkcinio stiliaus tekstynas
- PKA – pagrindinių komponentų analizė
- PST – publicistinio funkcinio stiliaus tekstynas
- TK – teminės koncentracijos įvertis
- VA – tarpveiksmažodiniai atstumai arba teksto sintaksinės struktūros kompleksiško įvertis
- VŽFI – vidutinis žodžio / žodžio formos ilgis

SAVOKŲ ŽODYNĖLIS

Aktyvumas (A) – tai kiekybinis teksto indikatorius, vertinantis teksto aktyvumą / dinamiškumą per būdvardžių ir veiksmažodžių santykį tekste (Zörnig ir Altmann, 2016).

Deskriptyvumas (D) – tai kiekybinis teksto indikatorius, vertinantis teksto deskriptyvumą / aprašomumą; yra priešingas aktyvumo indikatoriumi A (Kubát, Matlach ir Čech, 2014, 43; Kubát, 2016, 77–78).

Dispersinė analizė – duomenų analizė, kuri leidžia priklausomojo kintamojo skirtingas reikšmes paaiškinti vieno ar kelių nepriklausomų kintamųjų veikimu (Janilionis, Morkevičius ir Rauleckas 2008, 99).

Indikatorius a – kiekybinis teksto indikatorius, kuris yra normalizuotas h-taškas (žr. žemiau). Žymi labai dažnų žodžių ir žodžių formų srities dydį tekste (Zörnig, Kelih ir Fuks, 2016). Kuo a vertė didesnė, tuo analizuojamo teksto mažesnę dalį sudaro labai dažni žodžiai ir žodžių formos, o kuo šio indikatoriaus vertė mažesnė – tuo, priešingai, labai dažni žodžiai ir žodžių formos sudaro didesnę dalį teksto.

Indikatorius R_1 – kiekybinis teksto indikatorius, žymintis žodžių ir žodžių formų (retesnių) įvairovę (Zörnig, Kelih ir Fuks, 2016; Xiao ir Sun, 2018). Kuo didesnė indikatoriaus vertė (galimos vertės pasiskirsto nuo 0 iki 1), tuo analizuojamam tekstui būdingesnė pasikartojimų atžvilgiu retesnių žodžių ir žodžių formų įvairovė.

Indikatorius RR_{mc} – kiekybinis teksto indikatorius, žymintis teksto žodyno koncentraciją (Popescu, 2009, 166). Kuo tekste yra mažesnis labai dažnų žodžių ir žodžių formų skaičius su itin dideliais pasikartojimų dažniais, tuo teksto žodynas yra koncentruotesnis. Disertacijoje atliekamo tyrimo atžvilgiu šį indikatorių tikslingiau laikyti žodžių ir jų formų koncentracijos tekste įverčiu, nes indikatoriaus verčių skaičiavimui naudojamos žodžių formos, o ne lemos.

h-taškas – žymi žodžio / žodžio formos poziciją dažninėje lentelėje, kur žodžio / žodžio formos dažnis yra lygus jo rangui (pozicijai) toje dažninėje lentelėje (Popescu, 2009, 17). Kalbos analizės požiūriu, h-taškas žymi apytikslę ribą tarp labai dažnų (paprastai – funkcinų) žodžių ir žodžių formų ir retesnių (teminių / turinio, t. y. priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų.

Pagrindinių komponentų analizė (PKA) – tai duomenų analizės metodas, skirtas daugiamačių duomenų dimensijų mažinimui, išlaikant duomenų reprezentatyvumą (Smith, 2002).

Tarpveiksmožodiniai atstumai (VA) – tai teksto sintaksinės struktūros kompleksiskumo įvertis, kuris vertina tekste vidutinį atstumą žodžiais ir žodžių formomis tarp 2 iš eilės einančių veiksmažodžių / veiksmažodžių formų (Kubát, Matlach ir Čech, 2014, 42).

Kvartilinis plotis – tai duomenų sklaidos charakteristika, naudojama siekiant įvertinti, duomenų išsisklaidymą jų vidurkio atžvilgiu (Upton ir Cook, 1996, 55); dažnai naudojama išskirčių (netipiškų duomenų verčių) radimui.

Teksto dažninė struktūra – tai žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu & Altmann, 2015).

Teminė koncentracija – tai kiekybinis indikatorius, vertinantis iš esmės retesnių (priklausančių pagrindinėms kalbos dalims), tačiau konkrečiame tekste dažnumu labai dažniems (paprastai – funkciniais) beveik prilygstančių, žodžių ir žodžių formų kiekį tekste. Šie turinio žodžiai ir žodžių formos yra savotiška anomalija, todėl laikoma, kad jie nusako teksto temą (Kubát, Matlach ir Čech, 2014, 48–49). Kuo indikatorius vertė didesnė, tuo tekstas yra temiška centruotesnis.

Vidutinis žodžio / žodžio formos ilgis (VŽFI) – tai kiekybinis indikatorius, dažnai naudojamas tekstų sudėtingumui suprantamumo prasme (angl. *readability*) įvertinti (van Oosten, Tanghe ir Hoste, 2010).

ĮVADAS

Temos aktualumas ir naujumas

XX a. 3–4 dešimtmetyje Prahos lingvistinės mokyklos mokslininkų, analizavusių čekų bendrinės kalbos vartosenos diferenciaciją, buvo pastebėta, kad kalbos vartojimo ypatumai priklauso nuo situacijos ir funkcijų, kurias ji toje situacijoje atlieka. Tad bendrinės kalbos atmainos, kurias skiria atliekamos funkcijos, susijusios su kalbėjimo situacija arba, kitaip, vartojimo sritimi, buvo pavadintos funkciniais stiliais (Daneš, 1987, 13). Buvo pradėti tyrimai, kuriuose analizuotos atskirų funkcinių stilių (toliau – FS) kalbos priemonės, bandyta rasti ryšį tarp jų pasirinkimo bei vartojimo, kombinavimo, kalbinių ir nekalbinių veiksnių. Lietuvos mokslininkai – Juozas Pikčilingis (1971a, 1971b), Kazimieras Župerka (1983; 1995; 2000; 2004; 2016), Pranas Kniūkšta (2003; 2005), Regina Koženiausienė (1999; 2007; 2011; 2013a; 2013b), Rasuolė Vladarskienė (2000a; 2000b; 2002; 2004; 2005; 2006; 2007; 2012; 2014), Ona Petrėnienė (2003; 2011; 2013), Irena Smetonienė (2000; 2001; 2003; 2005; 2009; 2013a; 2013b; 2013c), Audronė Bitinienė (1983; 1997; 1998; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007a; 2007b; 2007c), Vida Žilinskienė (2001; 2002a; 2002b; 2002c; 2003a; 2003b; 2005; 2010), Aurelija Leonavičienė (2004; 2005; 2007; 2010) ir kiti – perėmė ir plėtojo šį funkcinės stilistikos požiūrį.

FS galima apibrėžti kaip kalbos posistemės, kurioms būdingi tam tikri požymiai (leksiniai, sintaksiniai, kiti) ir kurios yra susijusios su skirtingomis visuomenės gyvenimo komunikacinėmis situacijomis (Havránek, 2014 [1964]; Galperin, 1977, 2—33; Chloupek, 1993a; Canning, 2017, 63—85). Šiandien FS tyrimai pasaulyje apskritai nėra itin populiarūs, mokslininkai labiau linkę tirti žanrus, pavyzdžiui, Shlomo Argamon ir kolegų (2003), Jussi Karlgren ir Douglass Cutting (1994), Miroslav Kubát (2016), Peter Muntigl ir Helmut Gruber (2005), Philipp Petrenz (2012), Dirk Snyman ir kolegų (2012), Efsthathios Stamatatos, Nikos Fakotakis ir George Kokkinakis (2000), Gerard Steen (1999), Cong Zhang ir Haitao Liu (2015) ir panašūs. Taip yra todėl, kad FS apima didelę labai skirtingų žanrų tekstų įvairovę, tad rasti bendrus juos skiriančius ir siejančius vardiklius yra gana sudėtinga. O žanrai apima mažesnes tekstų grupes, tad šiuo požiūriu rasti panašumus ar skirtumus yra paprasčiau. Kiekybiniai metodai, kurie taikomi šioje disertacijoje, leidžia ieškoti panašumų ir skirtumų FS, kurie subendrina įvairių žanrų tekstus, požiūriu tiriant didelį kiekį tekstinių duomenų, tad disertacijos autorė tęsia lietuvių kalbotyros FS tradiciją, įvertindama naujas kiekybinės lingvistikos galimybes. Kitaip tariant, disertacijoje FS charakterizavimui pritaikomi nauji

metodai, o FS apibrėžimui ir išskyrimui remiamasi Lietuvos stilistikos mokyklos mokslininkų darbais.

FS ir jų savybės neretai tirtos ir tiriamos taikant kokybinius tyrimo metodus. Pavyzdžiui, Juozas Pikčilingis išskyrė 5 lietuvių kalbos FS (Pikčilingis, 1971a), Župerka tyrė kalbos priemonių konkurenciją (Župerka, 1995), metalingvistinės funkcijos raišką nelingvistiniuose tekstuose (Župerka, 2000), taip pat šis mokslininkas analizavo funkcinį stilių skyrimo kriterijus (Župerka, 2012), stiliaus normų savitumą (Župerka, 2016), bendrinės kalbos ir funkcinį stilių santykį (Župerka, 2005). Regina Koženiauskienė atskirai funkcinį stilių neanalizavo, tačiau jos darbai daug prisidėjo prie retorikos ir stilistikos mokslo tyrimų plėtotės Lietuvoje (Koženiauskienė, 1999; 2007; 2011; 2013a; 2013b). Taip pat paminėtini Irenos Smetonienės reklamos tekstų tyrimai (Smetonienė, 2000; 2001; 2005; 2013a; 2003b; 2013c), Rūtos Marcinkevičienės akademinį recenzijų tyrimai (2002), Daivos Pagojienės esė tyrimai (2006; 2009), Prano Kniūkštos (2003; 2005) bei Rasuolės Vladarskienės (2005; 2006; 2007) administracinės kalbos tyrimai ir pan.

Lietuvoje kiekybiniais metodais FS per sakinio ilgio ir struktūros parametrus tyrė Audronė Bitinienė (2001; 2002; 2004; 2007a; 2008; 2009; 2010). Ši mokslininkė, nors tyrė ir kitų FS sakinio ilgio ir struktūros parametrus, koncentravosi į publicistinį stilių. Statistinius metodus taip pat taiko Vida Žilinskienė, analizavusi gramatinių formų dažnumus ir pasiskirstymus FS (2002a; 2002b; 2002c; 2003; 2005; 2010). Paminėtina ir tai, kad kiekybiniais metodais mokslinio stiliaus mokslo populiarinamojo postilio tekstų kalbos ypatybes analizavo Ona Petrėnienė (2011). Vienas reikšmingų šių metodų taikymo darbų – Andriaus Utkos (2004) daktaro disertacija „Statistinis tekstų funkcijų nustatymas“, kuri yra susijusi su funkcinio požiūriu į kalbą, tačiau joje minėtas mokslininkas tiria ne FS, o tekstų funkcijas. Taigi matyti, kad Lietuvos mokslininkai pamažu imasi ir analizuoti stilius ir kiekybiniais metodais, tačiau reikia pasakyti, kad tokių tyrimų nėra daug.

Užsienyje FS statistinių tyrimų atlikta taip pat ne tiek ir daug. Kaip jau minėta, žanrų ir individualaus stiliaus tyrimai populiariesni nei FS analizė. Kita verta, galima paminėti Hannos Gnatchuk tyrimą, kuriame analizuotas samplaikų vartojimas anglų kalbos laikraštiname (angl. newspaper style) (vadinasi, publicistiniame) stiliuje (Gnatchuk, 2015). Taip pat minėtini Marco Guerini, Alberto Pepe ir Bruno Lepri darbai, kuriuose buvo tiriama kalbos priemonių vartojimo ir kalbos sudėtingumo įtaka mokslinių straipsnių (santraukų) populiarumui (Guerini, Pepe ir Lepri, 2012). Be to, Solomija N. Buk ir Andrij A. Rovenchak analizavo ukrainiečių kalbos FS (beletristinį (meninį), šnekamąjį, mokslinį, oficialųjį (verslo) ir publicistinį) per žodžių

dažnių ir jų pozicijų dažninėje lentelėje modelį (Buk ir Rovenchak, 2004). Tad, kaip matyti, FS tyrimų nėra gausu.

Kita vertus, kalbos (ir tekstų) tyrimų, kuriose taikomi kiekybinės lingvistikos (toliau – KL) metodai, esama nemažai. KL tyrimuose naudojami kiekybiniai duomenys, kiekybiškai įvertinami skirtingų kalbos lygmenų fenomenai ir modeliuojami jų tarpusavio ryšiai (Těšitelová, 1992, 13). Tai padeda geriau suprasti kalbos elementų ir procesų tarpusavio ryšius, iširti kalbos kitimo dinamiką ir pan. Taikydami kiekybinius metodus Radek Čech analizavo Čekijos prezidentų retoriką (Čech, 2011), Peter Zörnig ir Gabriel Altmann (2016) – Italijos prezidentų Naujųjų metų kalbų retoriką, Yu Fang ir Haitao Liu lygino teksto vertimų žodyno turtingumą (Fang ir Liu, 2015). Ioan-Iovitz Popescu, Radek Čech ir Gabriel Altmann, naudodami kiekybinę metodologiją, tyrė individualaus stiliaus raišką, žanrų požymius, vaikų kalbos ypatumus, patologinių tekstų kalbinę raišką (Popescu, Čech ir Altmann, 2011b). Popescu (2009) taip pat tyrė kiekybines žodžių savybes, kartu su Altmann – kalbos lygmenų diferenciaciją (Popescu ir Altmann, 2008). Altmann kartu su Reinhard Köhler analizavo įvairių tipų ir kalbos lygmenų pasikartojimus tekste (Altmann ir Köhler, 2015). Įvairūs kiekybiniai indikatoriai tyrėjų buvo naudojami tekstų kalbinės raiškos apibūdinimui ir įvairia logika parentų tekstų grupių atskyrimui, pavyzdžiui, tokius tyrimus atliko Ruina Chen ir Haitao Liu (2018), Chen, Liu ir Altmann (2016), Köhler ir Sven Naumann (2008), Miroslav Kubát (2016), Tayebah Mosavi Miangah ir Mohammad Javad Rezai (2016), Arjuna Tuzzi, Popescu ir Altmann (2010), Peter Zörnig, Emmerich Kelih ir Ladislav Fuks (2016) ir kiti.

Nors iš to, kas buvo rašyta, aiškėja, kad kalbos (ir tekstų) tyrimas taikant FS teorines prieigas pasaulyje nėra itin populiarus, tačiau, sekant Lietuvos stilistikos tradicija, šios disertacijos autorės buvo pasirinkta FS prieiga, nes kalbos vartojimas, siejamas su situacija ir kalbos atliekamomis funkcijomis, vis dėlto išlieka aktualus, kadangi visuomenėje vyksta nuolatiniai pokyčiai, kurie atsispindi ir kalbos vartojime. Lietuvių kalbos FS yra neblogai iširti kokybiniu aspektu, tačiau atliekant kokybinį tyrimą mokslininkas gali apdoroti tik ribotą tiriamosios medžiagos kiekį. Tad vienas iš naujų šioje disertacijoje dalykų – palyginti didelės apimties tiriamoji medžiaga – 3 tekstynai, kuriuos iš viso sudaro ~36,4 mln. žodžių ir 19 002 tekstai¹. Tokios

¹ Pavyzdžiui, V. Žilinskienė remiasi 300 000 žodžių tyrimo medžiaga (Žilinskienė, 2002b), A. Bitinienė – 50 000 žodžių apimties tyrimo medžiaga (Bitinienė, 2001), o, palyginimui, Andrius Utko teksto funkcijų nustatymo tyrime rėmėsi 10,68 mln. žodžių (463 tekstai) tyrimo medžiaga (Utko, 2004).

apimties medžiagos analizė gali padėti geriau atsakyti į klausimą, kokie požymiai būdingi administraciniam, moksliniam ir publicistiniam FS, bei ištirti tų požymių mastą.

Nors taikydami kokybinius metodus lietuvių mokslininkai identifikavo konkrečių FS požymius, ne visada yra iki galo aišku, kaip tie požymiai pasireiškia. Pavyzdžiui, kaip rašo Župerka, moksliniam stiliui būdingas apibendrinimas, objektyvumas, aiškumas (Župerka, 2012, 81), tačiau minėtas mokslininkas neaptaria, kokia yra šių požymių raiška ir koks šių požymių mastas. Taigi sudarant šiame tyrime naudotus tikslinius FS tekstynus disertacijos autorės buvo remiamasi ir kitų tyrėjų – pvz., Vladarskienės (2000a; 2004; 2005; 2006; 2007), Smetonienės (2001; 2005; 2009; 2013a; 2013b; 2013c) ir kitų – išskirtais FS požymiais, tačiau, kadangi nevisiškai aišku, kaip atrodo šių požymių raiška tekste, kaip ją įvertinti, šios disertacijos autorės buvo nuspręsta, kad disertacijoje analizuojamų FS kalbinė raiška bus tirama per kiekybinius požymius. Tad kitas naujas šioje disertacijoje aprašomo tyrimo bruožas yra toks: nors tyrime naudojami kiekybiniai kalbinę raišką žymintys indikatoriai rodo tik kai kuriuos kokybiniais metodais nustatytus kalbinės raiškos požymius, tačiau jie gali būti interpretuojami tiek matematiškai (paprasta suskaičiuoti, įvertinti mastą ir pasiskirstymą), tiek lingvistiškai (sudaro galimybes paaiškinti, ką skaitiniai rezultatai reiškia kalbos ar/arba teksto kontekste). Tad kiekybiniai indikatoriai sudaro prielaidas patikrinti FS kalbinės raiškos požymių mastą, pasiskirstymą, tipiskumą ir pan. didelės apimties tiriamojoje medžiagoje, kas leistų daryti apibendrinimus apie tiriamųjų FS kalbinę raišką ir jos ypatumus. Tai visiškai naujas žingsnis tiriant funkcinius lietuvių kalbos stilius.

Be to, atkreiptinas dėmesys į tai, kad statistiniuose FS tyrimuose, atliktuose, pavyzdžiui, tiek Bitinienės (1983; 1997; 1998; 2000; 2001; 2002; 2003; 2004; 2007a; 2007b; 2007c; 2010), tiek Žilinskienės (2001; 2002a; 2002b; 2002c; 2003a; 2003b; 2005; 2010), FS kalbinė raiška analizuota sakinio, o ne viso teksto lygmeniu², o šioje disertacijoje FS kalbinė raiška tirama teksto, o ne sakinio lygmeniu. Šio principo laikomasi, nes kalbinės raiškos požymiai tekstuose gali būti pasiskirstę netolygiai, tad mažai tikėtina, kad atsitiktinė tam tikro dydžio teksto atkarpa reprezentuotų, pavyzdžiui, monografijos ar romano raiškos ypatybes. Taigi ir šiuo požiūriu disertacijos tyrimas yra naujas.

² Pavyzdžiui, A. Bitinienė mokomųjų tekstų sakinio lygmens požymių tyrime rėmėsi 3100 sakinių stratifikuota imtimi (Bitinienė, 2007a).

Nors dauguma paminėtų darbų, kuriuose tirti FS, nepriklauso kiekybinės lingvistikos sričiai, tačiau nemažas darbų kiekis rodo, kad tema plačiai tyrinėjama. Kita vertus, kiekybinės lingvistikos metodų taikymas FS tirti yra nauja, nes iki šiol stilistikoje šie metodai nėra plačiai taikomi. Be to, skiriamųjų FS ypatybių tyrimas, siekiant jas geriau suprasti, gali prisidėti kuriant automatines tekstų paieškos ir klasifikavimo bei apskritai naujas natūralios kalbos apdorojimo sistemas ir atliekant tekstinių dokumentų kiekybinius tyrimus. Tai ypač aktualu šiais laikais, kai elektroninės informacijos kiekis pastoviai auga. Paminėtina, jog tai, kad indikatoriai gali būti naudojami automatiniam tekstų klasifikavimui į FS, parodė disertacijos autorės atliktas, bet į disertaciją netilpęs tyrimas, publikuotas kartu su bendraautoriais³. Be to, kai kurie statistiniai indikatoriai (*a*, TK, R₁, RR_{mc}, VA, etc.) lietuvių kalbai buvo pritaikyti pirmą kartą.

Taip pat galima sakyti, kad šios disertacijos tyrimas sudaro prielaidas FS tipologijos ir požymių papildymui bei empiriniam patvirtinimui/paneigimui. Nors, kaip jau buvo minėta, FS kokybiniu požiūriu yra neblogai ištirti, kiekybiniai metodai nebuvo labai aktyviai taikyti, tad ši disertacija – dar vienas žingsnis kiekybinių tyrimų linkme. Atliekant šioje disertacijoje aprašomą tyrimą, analizuojami ne tik atskiri kalbinės raiškos požymiai, bet ir jų kombinacijos, požymių tarpusavio ryšiai. Taip pat FS tiriama ne tik atskirai, bet analizuojami ir jų tarpusavio panašumai, skirtumai bei panašumų ir skirtumų mastas. Tad atveriamos naujos prieigos tekstų raiškos priemonių analizei bei sudaromos prielaidos patikrinti kokybiniais metodais išskirtus FS charakteristikas. Be to, disertacijoje lietuvių kalbai pritaikyti tyrimo būdai, patikrinta metodika gali leisti lietuvių kalbos tekstų tyrėjams toliau analizuoti ir skirtingų žanrų tekstų raiškos požymius.

Darbo objektas, tikslas ir uždaviniai

Šioje disertacijoje aprašomo tyrimo objektas yra 3 funkcinų stilių – publicistinio, mokslinio ir administracinio – rašytinių tekstų kalbinė raiška. Tikslas – kiekybinės lingvistikos metodais ištirti 3 lietuvių kalbos funkcinų stilių – administracinio, mokslinio ir publicistinio – kalbinę raišką ir įvertinti šių funkcinų stilių panašumus ir skirtumus pagal kalbinės raiškos požymių mastą. Šioje disertacijoje nėra siekiama kurti naujus metodus. Joje pritaikomi

³ Mandravickaitė, J., Krilavičius, T., & Man, K. L. (2019). Document Classification to Functional Styles (Domains of Use): Lithuanian Case. *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems*, 8(1), 38-41.

metodai lingvistinei problemai – minėtų FS stilių charakterizavimui – spręsti taikant kiekybinės lingvistikos metodus, t. y. taip, kaip ji anksčiau nebuvo sprendžiama. Šio darbo autorė, pasirėmusi Miroslav Kubát (2016), Peter Zörnig ir kolegų (2016), Ioan-Iovitz Popescu ir kolegų (Popescu et al., 2010; Popescu, Čech ir Altmann, 2011a; Popescu, et al., 2017), Radek Čech (2011; 2016), Čech ir kolegų (Čech, Popescu ir Altmann, 2014; Čech, Garabík, ir Altmann, 2015), Kubát ir Jiří Milička (2013) mokslininkų darbais ir sugretinusi kiekybinius požymius (indikatorius) su kokybinei prieigai skirtais požymiais, iškėlė prielaidą, kad FS kalbinei raiškai tirti, turėtų tikti šie indikatoriai:

- VŽFI – teksto sudėtingumo įvertis arba vidutinis žodžių ir žodžių formų ilgis tekste;
- a – labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų gausos tekste įvertis;
- R_1 – žodžių ir jų formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovės tekste įvertis;
- RR_{mc} – žodžių ir žodžių formų koncentruotumo įvertis – vertina žodžių ir žodžių formų, turinčių itin didelį dažnumą, kiekį tekste;
- MATTR – informacijos srauto įvertis arba vidutinis skirtingų žodžių ir žodžių formų kiekis tekste);
- TK – teminės koncentracijos įvertis – vertina iš esmės retesnių, pagrindinėms kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu bemaž prilygstančių funkciniams žodžiams ir žodžių formoms, kiekį tekste);
- A ir D – teksto „tono“ arba aktyvumo ir deskriptyvumo įverčiai – žymi veiksmažodžių ir būdvardžių santykį tekste;
- VA – teksto sintaksinės struktūros kompleksiško įvertis arba vidutinis atstumas tekste tarp 2 iš eilės einančių veiksmažodžių.

Plačiau apie šiuos indikatorius rašoma 1.4 sk.

Kad tikslas būtų pasiektas, kelti tokie uždaviniai:

1. Parengti 3 tikslinius tekstynus. Dabartinės lietuvių kalbos tekstyno⁴ pagrindu parengti administracinio ir mokslinio stiliaus tekstynus,

⁴ Paieška dabartinės lietuvių kalbos tekстыne prieinama adresu <http://tekstynas.vdu.lt/tekstynas/>

- delfi.lt tekstyno⁵ pagrindu – publicistinio stiliaus tekstyną. Apie tekstynų parengimą plačiau žr. 2 sk.
2. Parinkti metodiką kalbinei raiškai didelės apimties tekstynuose tirti kiekybiniu aspektu, įvertinant tiek atskiroms tekstų grupėms būdingus požymius, tiek tekstų grupių tarpusavio panašumus ir skirtumus pagal kiekybinių požymių raiškos mastą. Formuluojuojant išvadas pateikti rekomendacijų, į ką reikėtų atkreipti dėmesį.
 3. Išanalizuoti tiriamiems funkciniam stilium būdingą kalbinę raišką pagal 9 kiekybinius indikatorius – VŽFI, *a*, R₁, RR_{mc}, MATTR, TK, A, D ir VA – ir nustatyti, kas būdinga tiriamiems funkciniam stilium.
 4. Ištirti, kokia tiriamųjų požymių kombinacija būdingiausia konkrečiau FS kalbinei raiškai.
 5. Ištirti tikslinių funkcinių stilių kalbinės raiškos tarpusavio panašumus ir skirtumus požymių raiškos mastą ir atsakyti į klausimą, kuo tarpusavyje panašūs ir kuo – skiriasi tiriamieji funkciniai stiliai.
 6. Pateikti tekstų tyrėjams FS kalbinės raiškos modelius, t. y. susistemintus tikslinių funkcinių stilių „portretus“.

Tiriamoji medžiaga ir metodai

Tiriamoji **medžiaga** – Dabartinės lietuvių kalbos tekstyno⁶ pagrindu disertacijos autorės parengti administracinio ir mokslinio FS tekstynai bei projekto PASTOVU reikmėms sukurtas delfi.lt tekstyno⁷ pagrindu disertacijos autorės parengtas publicistinio FS tekstynas. Originalaus Dabartinės lietuvių kalbos tekstyno (DLKT) tekstams nebuvo galima taikyti disertacijoje numatytų tekstų analizės strategijų (plačiau žr. 2 sk.), todėl atitinkamos jo dalys (administraciniai ir moksliniai tekstai) buvo pritaikytos tyrimo reikmėms, o publicistinis tekstynas parengtas delfi.lt tekstyno pagrindu, nes publicistiniai tekstai, įtraukti į DLKT publikuoti tik iki 2008 m. imtinai. Publicistinis FS, fiksuodamas pokyčius visuomenėje, pats yra imlus

⁵ Projekto PASTOVU reikmėms sudarytas naujienų portalo Delfi straipsnių tekstynas, plačiau žr. <http://mwe.lt/>

⁶ Disertacijos autorė pateikė prašymą ir VDU Kompiuterinės lingvistikos centro vadovo dr. Andriaus Utkos leidimu galėjo disertacijos tikslams naudoti šį tekstyną.

⁷ Šis tekstynas viešai nėra prieinamas. Disertacijos autorė projekte PASTOVU dirbo jaunesniąja mokslo darbuotoja, todėl turėjo prieigą prie šio tekstyno. Kita vertus, viešai prieinami delfi.lt tekstyno pagrindu parengti dažniniai n-gramų sąrašai (<http://resursai.mwe.lt/atsisiusti>)

pokyčiams (Bitinienė, 2006; Butkutė, 2008; Leonavičienė, 2010), tad tikslinio publicistinio FS tekstyno pagrindu pasirinktas delfi.lt tekstynas, kurį sudaro 2014-2016 m. publikuoti tekstai. Tad, kaip jau minėta, šios disertacijos autorės tyrimui buvo parengti 3 tiksliniai tekstynai:

- administracinio FS tekstynas (toliau – AST), kurį sudarė apie 5,8 mln. žodžių, 4 527 tekstai;
- mokslinio FS tekstynas (toliau – MST), jame buvo apie 20,2 mln. žodžių, 1025 tekstai;
- publicistinio FS tekstynas (toliau – PST), kurį sudarė apie 10,4 mln. žodžių, 13 450 tekstų).

Nors lietuvių kalboje skiriami 5 FS (administracinis/dalykinis, mokslinis, publicistinis, meninis/beletristinis ir šnekamasis/buitinis), šioje disertacijoje apsispręsta tirti tik 3 FS – administracinio, mokslinio ir publicistinio – kalbinę raišką. Šnekamojo/buitinio stiliaus kalbinė raiška šioje disertacijoje neanalizuojama, nes tyrime apsiribota rašytinės kalbos atmaina, o surinkti didelės apimties buitinio stiliaus rašytinių tekstų tekstyną yra sudėtingas, daug laiko reikalaujantis darbas. Meninio/beletristinio stiliaus kalbinės raiškos apsispręsta netirti dėl apribojimų disertacijos apimčiai. Administracinis ir mokslinis FS turi tarpusavio panašumų tikslumo, aiškumo ir logiškumo požiūriu (Kniūkšta, 2005; Vladarskienė, 2007), tačiau juos vieną nuo kito skiria administracinio stiliaus standartiškumas/šabloniškumas (Žilinskienė, 2002b; Kniūkšta, 2005), tad tyrimui pasilikti abu. Publicistinis stilius atvaizduoja visuomenėje vykstančius įvykius bei daro poveikį visuomenės nuomonei (Bitinienė, 2007c, 12). Jį su moksliniu stiliumi sieja dalykiškumas, glaustumas, nuoseklumas, logiškumas, tačiau informacija dažnai pateikiama panašia į meninio stiliaus forma (Smetonienė, 2013). Tad į tyrimą pasirinkta įtraukti ir publicistinių stilių. Be to, pasakytina, kad nors šioje disertacijoje netiriama FS vidinė diferenciacija, tačiau į tikslinius tekstynus įtraukti visų postilių tekstų.

Šiame darbe taikomi kiekybinės lingvistikos (KL) **metodai** (plačiau žr. 1.3 skyrių). Visų pirma, sekant funkcinės stilistikos teorija, šios disertacijos autorės buvo pasiremta Pikčilingio (1971), Župerkos (1983; 1995; 2000; 2004; 2016), Bitinienės (1983; 1997; 1998; 2000; 2001; 2002; 2003; 2004; 2005; 2006; 2007a; 2007b; 2007c), Kniūkštos (2003; 2005), Vladarskienės (2000a; 2000b; 2002; 2004; 2005; 2006; 2007; 2012; 2014), Žilinskienės (2001; 2002a; 2002b; 2002c; 2003a; 2003b; 2005; 2010), Leonavičienės (2004; 2005; 2007; 2010), Petrėnienės (2003; 2011; 2013) išvalgomis, kad FS stiliai – administracinis, mokslinis ir publicistinis – turi jiems tipiškus,

skirtingų kalbinės raiškos požymių. Antra, išanalizavus KL mokslininkų darbus (žr. 1.3.3 sk.), buvo pasirinkti kiekybiniai indikatoriai, kurie, šios disertacijos autorės nuomone, akivaizdžiausiai apibūdina tiriamuosius FS. Tada buvo atlikta tiriamųjų FS kalbinės raiškos kiekybinė analizė. Tuomet buvo atliktas matematinis tyrimo rezultatų įvertinimas, skaitiniai rezultatai interpretuoti kalbos požiūriu (šiuo atveju – FS kontekste).

FS kalbinės raiškos kiekybinė analizė apėmė keletą etapų. Pirmame etape tiriamieji tekstiniai duomenys buvo paversti kiekybiniais duomenimis, t. y. kiekvienam kiekvieno tikslinio teksto tekstui buvo suskaičiuojami kalbinės raiškos požymius žymintys kiekybiniai indikatoriai. Antrame tyrimo etape atlikta aprašomoji duomenų analizė, kad būtų įvertintos pagrindinės statistinės indikatorių savybės tiriamųjų FS atžvilgiu (plačiau žr. 4.1.1, 4.2.1 ir 4.3.1 sk.). Tada buvo tiriami ryšiai tarp indikatorių kiekviename FS, t. y. atliekama koreliacinė analizė (plačiau žr. 4.1.2, 4.2.2 ir 4.3.2 sk.). Paskui taikyta pagrindinių komponentų analizė (PKA), kad būtų nustatytas kiekvieno tiriamojo FS kalbinę raišką geriausiai apibūdinančių indikatorių rinkinys (plačiau žr. 4.1.3, 4.2.3 ir 4.3.3 sk.). Paskutiniame etape buvo atlikta dispersinė analizė tiriamųjų FS kalbinės raiškos tarpusavio panašumų ir skirtumų nustatymui (plačiau žr. 5 sk.).

Darbe atliekant FS kalbinės raiškos analizę naudojamos šios kompiuterinės programos: *R*⁸, *QUITA*⁹, *MaWaTaTaRaD* ir *ThematicConcentration*¹⁰, morfologinis anotatorius¹¹. Programa *R* ir įvairios šiam statistiniam paketui skirtos bibliotekos naudotos tekstų sutvarkymui (plačiau žr. 2.2 sk.), pirminei analizei (plačiau žr. 3.1 sk.), indikatoriaus VA (plačiau žr. 1.4.3.3 sk.) verčių skaičiavimui, koreliacinei analizei (plačiau žr. 3.2 sk.), pagrindinių komponentų analizei (PKA) (plačiau žr. 3.3 sk.), dispersinei analizei (plačiau žr. 3.4 sk.), rezultatų vizualizavimui. Įrankis *QUITA* pasitelktas indikatorių VŽFI, *a*, *R*₁ ir *RR*_{mc} (plačiau žr. 1.4.3.1, 1.4.2.1-1.4.2.3 sk.) verčių skaičiavimui kiekvienam tiriamosios medžiagos tekstui. *MaWaTaTaRaD* naudotas indikatoriaus MATTR (plačiau žr. 1.4.2.4 sk.) vertėms kiekvienam tiriamų tekstynų tekstui suskaičiuoti. Teminės koncentracijos indikatoriaus (plačiau žr. 1.4.2.5 sk.) vertės buvo apskaičiuotos programa

⁸ Programą galima parsisiųsti adresu <https://www.r-project.org/>

⁹ Programą galima parsisiųsti adresu <https://miroslavkubat.webnode.cz/software/>

¹⁰ Programas galima parsisiųsti adresu <http://milicka.cz/en/software.htm>

¹¹ Morfologinis anotatorius buvo sukurtas projekto „Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas (PASTOVU)“ metu kaip pastoviųjų žodžių junginių atpažinimo įrankio dalis. Įrankis nėra viešai prieinamas. Disertacijos autorė projekte PASTOVU dirbo jaunesniąja mokslo darbuotoja, todėl turėjo prieigą prie šio įrankio.

ThematicConcentration. Visi tekstai buvo anotuoti morfologiškai (plačiau žr. 2.2 sk.), kad būtų galima vėliau rasti indikatorių A, D ir VA vertes (plačiau žr. 1.4.3.2-1.4.3.3 sk.).

Darbo struktūra

Disertaciją sudaro 5 skyriai, išvados ir priedai. Pirmas skyrius yra teorinio pobūdžio, jame pristatomos teorinės tyrimo priegigos, kurias sudaro 2 dalys: funkcinė stilistika (1.1-1.2 sk.) ir kiekybinė lingvistika (1.3-1.5 sk.). Antrame skyriuje apžvelgiama tyrimo medžiaga ir jos parengimas, trečiajame nuodugniai aprašoma tyrimo metodika ir eiga. Ketvirtajame skyriuje, kurį sudaro 3 poskyriai, išsamiai aprašomas kalbinės raiškos tyrimas kiekvienam tiriamųjų FS (administracinio, mokslinio, publicistinio) atskirai, o konkrečiau: aptariama aprašomoji (pirminė) duomenų analizė ir jos rezultatai, koreliacinė analizė ir PKA. Penktame skyriuje aptariami tiriamųjų FS kalbinės raiškos tarpusavio panašumai ir skirtumai, gauti taikant neparimetrinius dispersinės analizės metodus. Darbo pabaigoje teikiamos išvados, nurodomos tyrimo perspektyvos ir pridedami priedai – diagramos. Kiti priedai dėl didelės duomenų apimties pateikiami duomenų talpykloje¹². Disertacijoje yra 34 lentelės ir 17 paveikslų. Darbo pabaigoje pridedamas 256 pozicijų naudotos literatūros sąrašas.

Ginamieji teiginiai

Gynimui teikiami šie teiginiai:

1. Atskirų funkcinių stilių kalbinę raišką, nelyginant jų tarpusavyje, reprezentuoja skirtingos požymių kombinacijos.
2. Lyginant tikslinius funkcinis stilius tarpusavyje, kiekviename funkciniam stiliuje išryškėja skirtingas požymių raiškos mastas.
3. Kiekybiniai požymiai yra tarpusavyje susiję, t. y. vieno požymio raiškos kitimui turi įtakos kito požymio raiškos kitimas, todėl, analizuojant kalbinę raišką kiekybinės lingvistikos požiūriu, privalu tirti požymių, kuriuos žymi tyrime naudojami kiekybiniai indikatoriai, kombinacijas.

¹² Papildomi failai pasiekiami adresu <https://bitbucket.org/jstn00/papildomi-failai/downloads/>

4. Taikyta analizės metodika yra tinkama tikslinių funkcinių stilių modelių išskyrimui.

Darbo apibavimas

Doktorantūros metu paskelbti **straipsniai**:

1. Mandravickaitė Justina, Krilavičius Tomas ir Man Ka Lok (2019). Document Classification to Functional Styles (Domains of Use): Lithuanian Case. *International journal of design, analysis and tools for integrated circuits and systems (IJDATICS)*. Hong Kong: Solari Co, 2019, vol. 8, no. 1 (Proceedings of the International Conference on Recent Advancements in Computing in AI, IoT and Computer Engineering Technology (CICET 2019)).
2. Mandravickaitė Justina ir Krilavičius Tomas (2018). Quantitative Analysis of Textual Genres: Comparison of English and Lithuanian. *Proceedings of Int. Master and PhD students conference „Information Society and University Studies“ (IVUS 2018)*, p. 61—67.
3. Mandravickaitė Justina, Krilavičius Tomas ir Man Ka Lok (2018). Initial Analysis of Characteristics of Textual Genres: Comparison of Lithuanian and English. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 2.
4. Stanikūnas Daumantas, Mandravickaitė Justina ir Krilavičius Tomas (2017). Comparison of Distance and Similarity Measures for Stylometric Analysis of Lithuanian Texts. *Proceedings of Int. Master and PhD students conference „Information Society and University Studies“ (IVUS 2017)*, p. 1—7.
5. Mandravickaitė Justina ir Krilavičius Tomas (2017). Stylometric Analysis of Parliamentary Speeches: Gender Dimension. *Proceedings of the 6th workshop on BSNLP*, p. 102—107.
6. Mandravickaitė Justina ir Oakes Michael (2016). Multiword Expressions for Capturing Stylistic Variation Between Genders in the Lithuanian Parliament. *Proceedings of the LREC 2016 workshop „CCURL 2016 – Towards an Alliance for Digital Language Diversity“*, p. 80—85.
7. Mandravickaitė Justina ir Krilavičius Tomas (2015). Lietuvos Respublikos Seimo narių kalbinė raiška atsižvelgiant į jų politinę orientaciją. *Darbai ir Dienos* 64, p. 133—151.

Pranešimai konferencijose ir kituose mokslo renginiuose:

1. Mandravickaitė Justina ir Krilavičius Tomas. Similarities and Differences of Lithuanian Functional Styles: a Quantitative Perspective. *The 9th International Conference „Human Language Technologies – the Baltic Perspective“ (BALTIC-HLT 2020)*, 2020 09 22-23, Kaunas.
2. Mandravickaitė Justina ir Krilavičius Tomas. Document Classification to Domains of Use: Lithuanian Case. *Tarpt. konferencija „Duomenų analizės metodai programų sistemoms“ (DAMSS)*, 2019 11 28-30, Druskininkai.
3. Mandravickaitė Justina, Krilavičius Tomas ir Man Ka Lok. Document Classification to Functional Styles (Domains of Use): Lithuanian Case. *The International Conference on Recent Advancements in Computing in AI, IoT and Computer Engineering Technology (CICET 2019)*, 2019 10 21-23, Sinbėjus, Taivanas.
4. Mandravickaitė Justina. Kiekybiniai indikatoriai mokslinio ir publicistinio funkcinių stilių analizėje. *5-oji taikomosios kalbotyros konferencija „Kalbos ir žmonės: komunikacija daugiakalbiame pasaulyje“*, 2019 09 26-28, Vilnius.
5. Mandravickaitė Justina ir Krilavičius Tomas. Quantitative Indicators in the Analysis of Functional Styles. *Tarpt. konferencija „Duomenų analizės metodai programų sistemoms“ (DAMSS)*, 2018 11 29-12 01, Druskininkai.
6. Mandravickaitė Justina. Kiekybiniai indikatoriai rašytinių tekstų funkcinių stilių analizėje. *25-oji tarptautinė mokslinė Jono Jablonskio konferencija „Variantiškumas kalbose ir jų atmainose“*, 2018 09 27-28, Vilnius.
7. Mandravickaitė Justina ir Krilavičius Tomas. Quantitative Analysis of Textual Genres: Comparison of English and Lithuanian. *Int. Master and PhD students conference „Information Society and University Studies“ (IVUS 2018)*, 2018 04 27, Kaunas, Lithuania.
8. Mandravickaitė Justina, Krilavičius Tomas ir Man Ka Lok. Initial Analysis of Characteristics of Textual Genres: Comparison of Lithuanian and English. *International MultiConference of Engineers and Computer Scientists (IMECS)*, 2018 03 14-16, Honkongas.
9. Stanikūnas Daumantas, Mandravickaitė Justina ir Krilavičius Tomas. Comparison of Distance and Similarity Measures for Stylometric Analysis of Lithuanian Texts. *Int. Master and PhD students*

- conference „Information Society and University Studies“ (IVUS 2017), 2017 04 28, Kaunas.
10. Mandravickaitė Justina ir Krilavičius Tomas. Statistical Analysis of Word Frequency Distribution in Texts of Different Genres: Comparison of Lithuanian and English. *Tarpt. konferencija „Duomenų analizės metodai programų sistemoms“ (DAMSS)*, 2017 11 30-12 02, Druskininkai.
 11. Mandravickaitė Justina ir Krilavičius Tomas. Stylometric Analysis of Parliamentary Speeches: Gender Dimension. *The 6th Workshop on Balto-Slavic Natural Language Processing (BSNLP)*, 2017 04 04, Valensija, Ispanija.
 12. Mandravickaitė Justina ir Oakes Michael. Multiword Expressions for Capturing Stylistic Variation Between Genders in the Lithuanian Parliament. *LREC 2016 Workshop „CCURL 2016 – Towards an Alliance for Digital Language Diversity“*, 2016 05 23, Portorož, Slovėnija.
 13. Mandravickaitė Justina ir Krilavičius Tomas. Visualization of Differences in Language Usage of Parliamentary Position and Opposition in the Lithuanian Parliament 2008-2012. *Tarpt. Konferencija „Duomenų analizės metodai programų sistemoms“ (DAMSS)*, 2015 12 03-05, Druskininkai.

1. TEORINĖS TYRIMO PRIEIGOS

1.1 Funkcinė stilistika ir funkcinio stiliaus samprata

Kaip jau buvo minėta, šio darbo tikslas – taikant kiekybinės lingvistikos metodus ištirti administracinio, mokslinio ir publicistinio funkcinį stilių kalbinę raišką požymius ir jų masą, tad šiame skyriuje trumpai pristatomos funkcinės stilistikos teorinės prieigos ir FS samprata. Kalba, žvelgiant iš funkcijų pozicijos, laikoma sistema (kalbos priemonių, stilistinių ypatybių ir jų kombinacijų), skirta tam tikram tikslui pasiekti (Chovanec, 2014). Kalbos priemonių, susijusių su FS, vartojimas taip pat priklauso nuo teksto ar pasakymo funkcijos (Kontutyte, 2007). Tad FS galima apibūdinti kaip santykiškai nusistovėjusias raišką priemonių kombinacijas, skirtas bazinės komunikacijos funkcijos atlikimui (Chloupek, 1993b).

1.1.1 Prahos lingvistinė mokykla ir funkcinė stilistika: trumpos pastabos

Funkcinės stilistikos pagrindus XX a. 3–4 dešimtmetį padėjo Prahos lingvistinė mokykla – tyrėjai, analizavę vartosenos diferenciaciją čekų bendrinėje kalboje. Pasak Jan Chovanec (2014), tarp Prahos funkcinės stilistikos mokyklos atstovų Bohuslav Havránek (1893–1978) žinomiausias dėl savo darbų bendrinės kalbos ir FS srityse bei dėl požiūrio, kad kalbos taisyklingumas turėtų būti paremtas pasakymo funkcija, t. y. ar tai, kas ir kaip yra sakoma, leidžia pasiekti norimą komunikacinį tikslą, o ne istoriniais kriterijais, tokiais kaip kalbos grynumas (angl. *purity*).

Be Havránek, prie pagrindinių Prahos funkcinės stilistikos mokyklos atstovų, kurių darbai prisidėjo prie šios krypties susiformavimo, Chovanec (2014) taip pat priskiria Roman Jakobson (1896–1982), kuris išstobulino kalbos funkcijų komunikacijos akte koncepcijos supratimą. Taip pat išskiriamas Vilém Mathesius (1882–1945), savo darbuose pabrėžęs sinchroninės kalbos analizės svarbą bei domėjęs kalbos funkciniais aspektais visuose jos lygmenyse (Chovanec, 2014). Be to, Mathesius į funkcinę stilistiką įvedė elastingo stabilumo (angl. *elastic stability*) sąvoką kalbos kitimo ir kalbos kaip sistemos prisitaikymo prie pokyčių kontekste, atliko lyginamųjų tyrimų, susijusių su žodžių tvarkos sakiniuose problemomis anglų ir čekų kalbose (Mathesius, 1983, 138; Chovanec, 2014). O Jan Mukařovský (1891–1975) funkcinės stilistikos kontekste pristatė

automatizacijos (habitualizacijos) ir aktualizacijos sąvokas (Chovanec, 2014). Automatizacija jo buvo suprantama kaip kalbos priemonių vartojimas įprastais komunikuojantiems subjektams būdais, t. y. taip, kaip priimta ir tikimasi konkrečios kalbos vartotojų bendruomenės (Emmott ir Alexander, 2015, 290—291). Aktualizacija, minėto mokslininko nuomone, rodo kalbos priemonių vartojimą naujoviškai, kūrybiškai ar neįprastai (Doležel, 1968).

Kaip vienas kertinių tyrėjų Prahos lingvistikos mokyklos judėjime antroje XX a. pusėje išskiriamas ir Josef Vachek (1909–1996) (Chovanec, 2014). Savo darbuose jis išskėlė idėją, kad kiekvienos kalbos sistemos elementai yra glaudžiai tarpusavyje susiję, todėl daro poveikį vieni kitiems, o tai gali inicijuoti kalbos pokytį (Vachek, 1962). Vachek rašytinės kalbos tyrimai leido priėti prie išvados, kad rašytinė ir sakytinė kalbos atmainos veikia nepriklausomai pagal skirtingas funkcines normas (Vachek, 2014).

Šie ir kiti, nepaminėti¹³, Prahos lingvistikos mokyklos tyrėjai sukūrė ir išplėtojo FS koncepciją, kuri tebėra aktuali kalbos veikimo tyrimams. Šie tyrėjai į kalbą žvelgė kaip į sistemą, kuri kinta veikiami tiek išorinių (ekonominių, socialinių, kultūrinių ir kitų) veiksnių, tiek vidinių (kalbos sluoksniai, elementai ir struktūros, būdami tarpusavyje susiję, daro poveikį vienas kitam). Kalbą traktuodami kaip sistemą, kurios komponentai yra tarpusavyje susiję, Prahos lingvistikos mokyklos tyrėjai analizavo ne tik atskirus kalbos komponentus (lygmenis, struktūras ar elementus), bet ir jų tarpusavio ryšius bei vaidmenį visoje kalbos sistemoje. Be kita ko, šios mokyklos atstovų tyrimai buvo orientuoti į praktinius taikymus, ypač kalbų mokymą (Hall, 2014).

1.1.2 FS samprata

Funkcinės stilistikos darbuose paprastai tiriamos atskirų FS kalbos priemonės, bandoma rasti ryšį tarp jų pasirinkimo bei išdėstymo, kombinavimo ir nekalbinių veiksnių (Bitinienė, 2007a). FS yra atvira sistema. Šią sistemą galima tirti pagal stilių nulemiančius pragmatinius veiksnius bei nuo jų priklausančias atskiro stiliaus posistemes (postilius). Postiliams būdingi tam tikri uždaros sistemos požymiai, tačiau analizuojant konkrečius tekstus galima pastebėti, kad ribos tarp FS ir jų atmainų tampa sunkiau pastebimos, nes neretai įsiterpia kitiems stiliams būdingų elementų, pavyzdžiui, šnekamojo

¹³ Poskyryje pristatyti tik su disertacijos problematika susiję Prahos funkcinės stilistikos mokyklos atstovai. Daugiau žr. Chovanec (2014).

(buitinio) stiliaus elementų vartojimas publicistiniuose, meniniuose ar mokslo populiarinamuosiuose tekstuose (Smetonienė, 2013a).

Kadangi funkcinis požiūris į kalbą yra susijęs su kalbos vartojimo pragmatika, kalbos ir kalbėjimo kontekstas čia yra tiek pat svarbus, kiek ir formalūs požymiai, sudarantys kalbinį įvykį (Canning, 2017, 63–85). Maiklas Halliday (1971; 1994) dažnai laikomas funkcinės stilistikos pagrindinių koncepcijų autoriumi. Jis išskėlė idėją, kad kalbai būdingos trys svarbiausios funkcijos (ideacinė, tarpasmeninė ir tekstinė), kurių sankirta sukuria kalbėjimo ar teksto prasmę. Ideacinė funkcija yra reprezentacijos priemonė, tarpasmeninė – transakcijos, tekstinė – informacijos arba pranešimo priemonė (Canning, 2017, 63–85; Coupland, 2007, 12–13). Sekant šiuo funkcinio požiūriu, FS yra būdas, kuriuo tekstas reaguoja į funkcinio objekto poreikius (Daneš, 1987, 13), t. y. tai, koks FS pasitelkiamas tam tikroje situacijoje, priklauso nuo (komunikacinio) tikslo – ko siekiama tam tikra interakcija, pavyzdžiui, informuoti, įtikinti ar kita.

Jau minėtas Prahos lingvistinės mokyklos narys Havránek (2014 [1964]) stilių apibrėžė kaip kalbinės raiškos priemonių organizaciją (t. y. kaip jos išdėstomos, kombinuojamos tarpusavyje), o Mathesius nuomone, stilius apibūdina manierą ar būdą, kuriuo kalbinės priemonės vartojamos konkrečiam tikslui pasiekti (Hausenblas, 1994, 313). Tad stilių galima traktuoti kaip užbaigto teksto (tai gali būti bet kas nuo pasakymo iki meno kūrinio) savybę arba kaip tam tikrą galimybę, nulemtą teksto situacijos, kurią sudaro tiek kalbinė medžiaga, tiek kalbėtojo ar rašytojo individualumas, tiek kalbėjimo tikslas (Daneš, 1994). Pirmu atveju turime individualųjį stilių, antruju – FS. Havránek skiria FS pagal juose vyraujančias kalbos funkcijas, t. y. komunikacinė funkcija vyrauja šnekamajame stiliuje, praktinė techninė (angl. workaday technical), kitaip – praktinė specialioji funkcija (Vladarskienė, 2000a) – dalykiniame stiliuje, teorinė techninė (angl. theoretical technical) arba teorinė specialioji funkcija (Vladarskienė, 2000a) – moksliniame stiliuje, o estetinė – grožinės literatūros stiliuje. Tad Prahos lingvistinėje mokyklos FS sampratoje dominuojanti vaidmenį atlieka kalbos funkcijos, nusakančios komunikacinį tikslą.

Rusų mokslininkas Ilja R. Galperin laikosi kiek kitokios pozicijos. Nors, anot Galperin, FS yra kalbos priemonių sistema, leidžianti pasiekti tam tikrą komunikacinį tikslą (Galperin, 1977, 249), tačiau, tirdamas anglų kalbą, FS jis išskiria remdamasis ne tik kalbos funkcijų, nusakančių interakcijos tikslą, kriterijumi. Anglų kalboje Galperin išskiria 5 FS: beletristinį (meninį), publicistinį (išsiskiria įtikinėjimu ir poveikiu visuomenės nuomonei (Galperin, 1977, 287)), laikraščių (žurnalistinį) (skiriamieji požymiai – informatyvumas ir vertinimas) (Galperin, 1977, 297)), mokslinės prozos

(mokslinį), oficialiųjų dokumentų (administracinį). Šis tyrėjas pripažįsta, kad vertinamuoju aspektu laikraščių (žurnalistinis) stilius turi panašumų su publicistiniu stiliumi, tačiau laikraščių tekstams būdinga nusistovėjusi informacijos pateikimo struktūra, specifinis žodynas bei vartojamos sintaksinės konstrukcijos, jo manymu, leidžia išskirti savarankišką laikraščių funkcinį stilių (Galperin, 1977, 297). Kitaip sakant, Galperin išskirti publicistinis ir laikraščių FS vienas nuo kito skiriasi pagal kalbinius požymius (kalbines ypatybes ir stilistines priemones).

Apibendrinant galima teigti, kad Havránek FS sampratoje dominuojanti vaidmenį atlieka kalbos funkcijos (t. y. nekalbiniai FS skyrimo kriterijai). Kita vertus, Galperin FS sampratoje kalbiniai kriterijai (kalbinės ypatybės ir stilistinės priemonės) laikomi tiek pat svarbiais, kiek ir nekalbiniai. Lietuvos stilistikos mokyklos tyrėjai apibrėždami FS laikosi pirmojo požiūrio, kuris detaliau pristatomas kitame poskyryje.

1.1.3 FS samprata Lietuvos stilistikos mokyklos tradicijoje

Lietuvių kalbos FS XX a. 7-am dešimtmetyje išskyrė Juozas Pikčilingis. Jo sudarytą lietuvių kalbos FS sistemą sudaro 5 stiliai: meninis (beletristinis), publicistinis, mokslinis, administracinis (oficialusis dalykinis) ir šnekamasis (buitinis) (Pikčilingis, 1971b). Šia sistema remiamasi ir šiandien. FS skaidomi į postilius, postiliai skyla į žanrus (Župerka, 2005). FS tiek Pikčilingis, tiek apskritai Lietuvos stilistikos mokyklos tyrėjai laiko bendrinės kalbos atmainą, kurios stilistines ypatybes ir kalbos priemones nulemia kalbos vartojimo sritis ir funkcijos (Pikčilingis, 1971b, 277; Župerka, 2005; Župerka, 2012, 78). Kitaip sakant, Lietuvos stilistikos mokyklos tyrėjų FS sampratoje, kaip ir Havránek suformuluotoje (žr. 1.1.2 sk.), dominuoja nekalbiniai kriterijai. Pikčilingis FS skiria pagal 3 kalbos funkcijas: komunikacinę, informacinę ir poveikio (Pikčilingis, 1971b, 280). Župerka taip pat išskiria 3 kalbos funkcijas, jas įvardindamas kaip bendravimo, pranešimo ir poveikio (Župerka, 2012, 79). Pastebėtina, kad Pikčilingio komunikacinę funkciją Župerka vadina bendravimo, o informacinę – pranešimo, tačiau esmė lieka ta pati. Taigi galima tarti, kad lietuvių stilistai kalbos funkcijas suvokia vienodai.

Župerka, reprezentacinę kalbos funkciją siedamas su kalbos aktu, ją vadina dalykinio pranešimo (Župerka, 2012, 99), o referentinę – komunikacine (Župerka, 2012, 79). Komunikacinę funkciją šis tyrėjas dar skirsto į kasdienio bendravimo, pranešimo ir poveikio. Lietuvių kalbos FS sistamai, anot Župerkos, dar svarbios yra ekspresinė (išreiškia autoriaus santykį tiek su teksto turiniu, tiek su jo adresantu), apeliacinė (kreipimosi į adresantą) ir

estetinė kalbos funkcijos (Župerka, 2012, 100), nors pastaroji yra aktuali ne kiekvienai komunikacinei situacijai. Apeliacinės funkcijos raiška taip pat yra ribota, nes ji taikoma ne visų FS tekstuose.

Apibendrinant galima sakyti, kad FS koncepcijoje, kuria iš esmės vadovaujasi ir Lietuvos tyrėjai, stilius yra suprantamas kaip tarnaujantis tam tikroms socialinėms-komunikacinėms funkcijoms (Grzybek, Stadlober, Kelih ir Antić, 2005). Tad FS yra susiję su tam tikromis diskurso sferomis, kitaip apibūdinamomis kaip kasdienis, oficialusis-administracinis, mokslinis, žurnalistinis ar meninis stiliai. Be to, FS yra susijęs su pragmatiniu komunikacijos aspektu. FS ribos nėra labai griežtos, nes stiliai nuolat kinta. Vieną stilių nuo kito skiria kalbos ir stilistinių priemonių vartojimo ypatumai – kalbos priemonių atranka, jų kombinacijos ir kt. Šiame darbe, sekant Lietuvos stilistikos mokyklos tradicija ir Župerkos (2012, 78) apibrėžimu, FS laikoma bendrinės kalbos atmaina, kurios raišką (kalbines priemones ir stilistines ypatybes) lemia kalbos vartojimo sritis, atliekamos funkcijos bei turinys.

1.2 Funkcinių stilių sistema ir jų požymiai

1.2.1 FS tipologijos

Kaip jau minėta ankstesniame skyriuje, FS galima apibrėžti kaip kalbos posistemes, kurioms būdingi tam tikri požymiai (leksiniai, sintaksiniai, kiti) ir kurių taikymas yra susijęs su skirtingomis komunikacinėmis situacijomis visuomenės gyvenimo srityse. Kadangi tyrėjai jas apibrėžia ir skiria nevienodai, tai ir išskiriamų FS skaičius gali būti skirtingas. Šie skirtumai priklauso tiek nuo egzistuojančių konkrečios kalbos bendrinės kalbos ištirtumo lygmens, tiek nuo stilių skyrimo kriterijų. Stilių gali daugėti kalbai sluoksniuojantis pagal vartosenos sferas arba mažėja stiliams panašėjant (Župerka, 2005). Tad galima ne viena FS tipologija. Šiame skyriuje pateikiami keli FS tipologijos pavyzdžiai, sudaryti rusų (Kuznetz ir Skrebnev, Arnold, Galperin), čekų (Havránek) ir lietuvių (Pikčilingio ir Župerkos) autorių (žr. 1 lentelę).

1 lentelė. FS tipologijų pavyzdžiai

Kuznetz ir Skrebnev (1960, cituota iš Hernández-Campoy, 2016, 27)	Arnold (1981, cituota iš Hernández-Campoy, 2016, 27)	Galperin (1977)	Havránek, 2014 [1964], 37)	Pikčilingis (1971a, 1971b)	Župerka (2012, 80—97)
1. Literatūrinis (knyginis) stilius	1. Literatūriniai-knyginiai stiliai	1. Beletristinis (meninis) stilius	1. Šnekamasis stilius (angl. <i>conversational</i>)	1. Oficialusis dalykinis stilius	1. Administracinis stilius
publicistinis stilius	mokslinis	poezija	2. Dalykinis stilius (angl. <i>workaday (matter-of-fact)</i>)	juridinis postilis	a) juridinis postilis
mokslinis (technologinis) stilius	oficialiųjų dokumentų	emocionali proza	3. Mokslinis stilius	kanceliarinis postilis	b) kanceliarinis postilis
oficialiųjų dokumentų stilius	publicistinis (laikraščių)	drama	4. Grožinės literatūros stilius (angl. <i>poetic</i>)	diplomatinis postilis	c) diplomatinis postilis
2. Laisvasis (šnekamasis) stilius	retorinis	2. Publicistinis stilius		2. Mokslinis stilius	2. Mokslinis stilius
literatūrinis šnekamasis stilius	poetinis	viešos kalbos		mokslinis postilis	teorinis (specialusis) postilis
intymusis šnekamasis stilius	2. Šnekamieji stiliai	esė		mokslo populiarinamasis postilis	mokomasis postilis
	literatūrinis šnekamasis	straipsniai		3. Publicistinis stilius	3. Publicistinis stilius
	intymusis šnekamasis	3. Laikraščių stilius		laikraščių postilis	informacinis postilis
	kasdienis šnekamasis	žinutės		publicistikos (siauraja prasme) postilis	analitinis postilis

		antraštės			ekspresyvusis postilis
		reklama ir skelbimai		4. Kasdienis buitinis (šnekamasis) stilius	4. Šnekamasis stilius
		redakciniai straipsniai		5. Meninis (beletristinis) stilius	5. Meninis stilius
		4. Mokslinės prozos stilius		poezija	
		tikslųjų mokslų postilis		proza	
		humanitarinių mokslų postilis			
		mokslo populiari-namasis postilis			
		5. Oficialių- jų dokumen- tų stilius			
		verslo dokumen- tai			
		teisiniai dokumen- tai			
		diplomati- nė kalba			
		kariniai dokumen- tai			

1 lentelėje pateiktos FS tipologijų, su kuriomis disertacijos autorė susipažinusi išsamiausiai, pavyzdžiai. Pavyzdžiui, M. D. Kuznetz ir Y. M. Skrebnev (1960, cit. iš Hernández-Campoy, 2016, 27) FS sistemą sudaro 2 FS – literatūrinis (knyginis) ir šnekamasis. Iš esmės autoriai skiria rašytinę ir sakytinę kalbos atmainas jas įvardindami kaip dideles FS grupes, kurias vėliau skaido į savarankiškus stilius, o ne į postilius. Literatūrinį (knyginį) FS Kuznetz ir Skrebnev dar skirsto į publicistinį, mokslinį (technologinį) ir

oficialiųjų dokumentų stilius, o laisvąjį (šnekamąjį) stilių – į literatūrinį šnekamąjį ir intymųjį šnekamąjį. Irinos V. Arnold (1981, cituota iš Hernández-Campoy, 2016, 27) sudarytą FS sistemą, kaip ir Kuznetz ir Skrebnev (1960, cit. iš Hernández-Campoy, 2016, 27) sudaro 2 didelės FS grupės (literatūriniai-knyginiai ir šnekamieji), tačiau detalesnis šių grupių skirstymas į FS skiriasi.

Kaip galima matyti 1-oje lentelėje, Arnold literatūrinių-knyginių stilių grupėje išskiria ne tik publicistinį (laikraščių), mokslinį ir oficialiųjų dokumentų FS, bet ir retorinį bei poetinį. Šnekamųjų stilių grupėje Arnold išskiria 2 tuos pačius stilius kaip ir Kuznetz ir Skrebnev – literatūrinį šnekamąjį ir intymųjį šnekamąjį, bei prideda trečiąjį stilių – kasdienį šnekamąjį. Paminėtina, kad Arnold, išskirdama FS grupes, priešingai nei Kuznetz ir Skrebnev, vartoja daugiskaitą – išskiria literatūrinius-knyginius stilius, o ne stilių, bei šnekamuosius stilius, o ne šnekamąjį stilių. FS stilių grupes, kaip ir Kuznetz ir Skrebnev, Arnold taip pat skirsto į savarankiškus stilius, o ne į postilius. Tad galima sakyti, kad šias 2 FS tipologijas sudaro savarankiški FS, į stambesnes FS grupes grupuojami pagal kalbos atmainą – rašytinę arba sakytinę.

Galperin (1977) skiria 5 FS – beletristinį (meninį), publicistinį, laikraščių, mokslinės prozos ir oficialiųjų dokumentų. Priešingai nei pirmosios 2 aptartos tipologijos, skirstymas neprasideda nuo kalbos atmainų (didelių FS grupių), o iš karto išskiriami 5 FS pagal kalbos vartojimo sritį. Galperin FS skirstymas į postilius remiasi nevienoda logika. Meninis (beletristinis) stilius skirstomas į poeziją, emocionaliąją prozą ir dramą – remiamasi literatūrologijoje įprastu grožinės literatūros skirstymu. Publicistinį stilių Galperin skaido į viešas kalbas, esė ir straipsnius – remiamasi mišria logika. Viešas kalbas nuo esė ir straipsnių skiria tai, kad jos priklauso sakytinei kalbos atmainai. Esė ir straipsniai išskirti remiantis žanrine diferenciacija.

Paminėtina, kad Galperin FS klasifikacija skirta anglų kalbai, o esė priskyrimas publicistiniam stiliui priklauso anglų kalbos žanrų tradicijai (Pagojienė, 2009, 5). Šio stilisto laikraščių stilius, kaip minėta 1.1. sk., iš dalies panašus į publicistinį stilių, tačiau laikraščių tekstams būdinga nusistovėjusi informacijos pateikimo struktūra, specifinis žodynas bei kalbinės raiškos priemonės, todėl, Galperin manymu, galima išskirti savarankišką laikraščių stilių (Galperin, 1977, 297). Šį stilių Galperin dar skirsto į žinutes, antraštes, reklamą ir skelbimus, redakcinius straipsnius. Tad galima sakyti, kad laikraščių stilių Galperin skirsto į žanrus, o ne į postilius.

Mokslinės prozos (mokslinį) stilių šis stilistas skirsto pagal mokslo sritis ir išskiria tikslųjų mokslų postilį bei humanitarinių mokslų postilį. Taip pat atskiru laikomas mokslo populiarinamasis postilis, kuris skiriamas todėl, kad

jo tikslas – ne tik informuoti, pateikti naujas žinias, kaip pirmuosiuose mokslinės prozos postiliuose, bet ir tai daryti patraukliai, įsimintinai, t. y. naudojamos smarkiai besiskiriančios stilistinės ir kalbinės priemonės, ką Galperin savo FS tipologijoje yra labiau linkęs pabrėžti nei atliekamas skirtingas kalbos funkcijas. Galiausiai, oficialių dokumentų (atitinka administracinį) stilių šis tyrėjas skirsto į verslo dokumentų, teisinių dokumentų, diplomatinės kalbos ir karinių dokumentų postilius, vėl pasiremdamas žanrine tipologija. Paminėtina, kad Galperin, skirtingai nei minėtose Kuznetz ir Skrebnev, Arnold FS tipologijose, atskirai neišskiria šnekamojo stiliaus (išskyrus viešas kalbas, priskirtas publicistiniam stiliui).

Vieno svarbiausių Prahos funkcinės stilistikos mokyklos atstovų Havránek FS klasifikaciją sudaro 4 FS (funkcines kalbas) (Havránek, 2014 [1964], 37): šnekamasis, dalykinis, mokslinis ir grožinės literatūros. Jo klasifikacija remiasi kalbos funkcijomis: komunikacinė funkcija tipiška šnekamajam stiliui, praktinė techninė (angl. workaday technical), kitaip – praktinė specialioji funkcija (Vladarskienė, 2000a) būdinga dalykiniam stiliui, teorinė techninė (angl. theoretical technical) arba teorinė specialioji funkcija (Vladarskienė, 2000a) – moksliniam stiliui, o estetinė – grožinės literatūros stiliui. Havránek FS į postilius, t. y. smulkesnes FS kategorijas, neskirsto. Lietuviška FS tipologija yra artimesnė Havránek, Kuznetz ir Skrebnev bei Arnold FS tipologijoms nei Galperin.

Pikčilingio sudarytą lietuvių kalbos FS sistemą sudaro 5 stiliai: oficialusis dalykinis, mokslinis, publicistinis, kasdienis buitinis (šnekamasis) ir meninis (beletristinis). Šiuos FS Pikčilingis skiria pagal 3 kalbos funkcijas: komunikacinę, informacinę ir poveikio (Pikčilingis, 1971a, 280). Župerkos išskirti lietuvių kalbos FS sutampa su Pikčilingio, tik Župerka oficialųjį dalykinį stilių vadina administraciniu. Kita vertus, skiriasi vidinė Pikčilingio ir Župerkos FS diferenciacija. Oficialiojo dalykinio arba administracinio stiliaus skaidymas į postilius sutampa – tiek Pikčilingis, tiek Župerka šį stilių skirsto į juridinį, kanceliarinį ir diplomatinį postilius. Mokslinį stilių Pikčilingis skirsto į mokslinį ir mokslo populiarinamąjį postilius, o Župerka – į teorinį (specialųjį) ir mokomąjį. Pikčilingio mokslinį postilį atitinka Župerkos teorinis (specialusis) postilis. Mokslo populiarinamąjį postilį Župerka laiko teorinio (specialiojo) postilio dalimi (Župerka, 2012, 80), o Pikčilingis mokomojo postilio neišskiria.

Publicistinių stilių abu mokslininkai į postilius skaido skirtingai: Pikčilingis išskiria laikraščių ir publicistikos (siaurąją prasme) postilius, o Župerka – informacinį, analitinį ir ekspresyvių postilius. Jo vidinė publicistinio stiliaus diferenciacija remiasi skirtingu pranešimo ir poveikio funkcijų santykiu (informaciniam postiliui būdinga pranešimo funkcija, analitiniam – pranešimo

ir poveikio funkcijos yra vienodai svarbios, o ekspresyviajame postilyje vyrauja poveikio funkcija). Pastebėtina, kad, priešingai nei kitose 1 lentelėje pateiktose FS tipologijose, Havránek FS tipologijoje publicistinis stilius nėra išskirtas. Kasdienio buitinio (šnekamojo) stiliaus Pikčilingis ir Župerka, priešingai nei Kuznetz ir Skrebnev bei Arnold, į postilius neskirsto. Galiausiai, meninio (beletristinio) stiliaus Župerka detaliau į postilius neskirsto, o Pikčilingis išskiria poezijos ir prozos atmainas, panašiai kaip Galperin remdamasis literatūrologijai būdingu grožinės literatūros skirstymu.

Kaip galima matyti šiame poskyryje, nėra universalios FS tipologijos, tad šios disertacijos autorė remiasi lietuviška FS skyrimo tradicija. Kadangi nevienodai žiūrima į FS, pateiktais 1 lentelėje FS tipologijų pavyzdžiais norima parodyti pasirinktos tipologijos panašumus ir skirtumus kitų FS tipologijų kontekste. Tolesniuose poskyriuose pristatomi disertacijoje tiriami lietuvių kalbos FS – administracinis, mokslinis ir publicistinis. Šnekamasis ir meninis FS neanalizuojami, pirmasis – kadangi priklauso sakybinei kalbos atmainai, o antrasis – dėl ribojimų disertacijos apimčiai.

1.2.2 Lietuvių kalbos administracinio stiliaus ypatybės

Nors tikėtina, kad panašūs požymiai būdingi ir kitų kalbų administraciniams stiliams, tačiau disertacijos tikslas, kaip jau minėta, yra ištirti lietuvių kalbos FS, tad šiame poskyryje apžvelgiami tik lietuvių mokslininkų darbai, kuriuose tiriamas lietuvių kalbos administracinis stilius. Administracinio stiliaus paskirtis – perteikti santykius tarp visuomenės narių bei tarp visuomenės narių ir valstybės (Gruodytė, 1994). Vykstant oficialiam bendravimui, perduodamu pranešimu siekiama paveikti adresatą kalbos priemonėmis. Administracinio stiliaus tekstuose, visų pirma, siekiama išplėsti / papildyti adresato turimą informaciją (vykdoma pranešimo funkcija), o antra – daryti poveikį adresato elgesiui (vykdoma direktyvinė arba valdymo funkcija). Taigi matyti, kad, kaip rašo Vladarskienė, administracinio stiliaus tekstuose dominuoja dvi funkcijos, kurios yra glaudžiai susijusios (Vladarskienė, 2004, 24). Tą patvirtinta ir kiti tyrėjai. Pavyzdžiui, anot Župerkos, administraciniame stiliuje vyrauja pranešimo funkcija, o šalia jos pastebima ir valdymo (poveikio) funkcija, nes valstybės piliečiai įstatymų laikosi, o nutarimus (vyriausybės ir panašiai) – vykdo (Župerka 2012, 85). Administracinis stilius dar vadinamas dalykiniu (Žilinskienė, 2002c).

Kita vertus, Župerka pastebi, kad kai kurių administracinio stiliaus postilių tekstuose (pvz., prašymuose, oficialiuose laiškuose ir pan.) išryškėja apeliacinė funkcija (Župerka, 1983, 108). Andriaus Utkos atlikto kiekybinio

tekstų funkcijų nustatymo rezultatai parodė, kad nors administracinio stiliaus tekstuose (oficialiuosiuose dokumentuose) direktyvinė / valdymo funkcija dominuoja, tačiau apeliatyvumo funkcija juose nebuvo išreikšta, t. y. šiuo atžvilgiu administracinio stiliaus tekstai šiai funkcijai buvo neprototipiškiausi (Utkā, 2004, 106-107). Tad kokybinio tyrimo išvalga atlikus tyrimą su kiekybiniais duomenimis šiuo atveju nepasitvirtino.

Anot Vladarskienės (2000), administracinio stiliaus postilius galima skirti pagal tris nekalbinius veiksnius: 1) skirtingą kalbos vartojimo sritį, 2) skirtingą pranešimo tikslą (pranešti, norminti ir panašiai) ir 3) skirtingą santykį tarp adresanto ir adresato. Lietuvių tyrėjų darbuose paprastai išskiriami šie administracinio stiliaus postiliai: valstybės dokumentų (Vladarskienė, 2004, 24), kitaip – juridinis (Bitinienė, 1998; Župerka, 2012, 85), kanceliarinis (apima, pavyzdžiui, aktus, įsakymus, protokolus, įgaliojimus ir pan.) bei diplomatinis (apima, pavyzdžiui, deklaracijas, konvencijas, sutartis ir panašiai) (Bitinienė, 1998; Vladarskienė, 2004, 25; Župerka, 2012, 85). Šioje disertacijoje į postilius nėra atsižvelgiama ir vidinė FS diferenciacija nėra tiriama, tačiau tikslinį administracinio funkcinio stiliaus tekstyną AST sudaro visų 3 postilių tekstai (plačiau žr. 2.3.1 sk.).

Pranas Kniūkšta administracinę kalbą vadina įvairių dokumentų kalba (Kniūkšta, 2005). Anot šio tyrėjo, administracinis stilius yra gana sustabarėjęs, kalbos naujovės išleidžiamos tik specialioms sąvokoms ir reiškiniams nusakyti. Kaip teigia Vladarskienė ir Župerka, administraciniam stiliui būdingas oficialumas (Vladarskienė, 2004, 26; Župerka, 2012, 85). Tai lemia stiliaus vartojimo sfera, t. y. bendravimas oficialiose situacijose. Minėti tyrėjai taip pat teigia, kad standartiškumas taip pat laikomas skiriamuoju šio stiliaus požymiu (Vladarskienė, 2004, 26; Župerka, 2012, 85). Šiam stiliui būdingos standartinės dokumentų pateikimo formos, t. y. šablonai, šabloniškos frazės ir pan. Administracinio stiliaus standartiškumas pašalina dviprasmybių galimybę ir sukuria sąlygas glaustumui. Dar viena šio stiliaus ypatybė yra dalykinis tikslumas (Pikčilingis, 1971a; Bitinienė, 1998; Vladarskienė, 2004, 26; Župerka, 2012, 85). Jis pasireiškia terminų vartojimu (Petrėnienė, 2003), dviprasmybių nebuvimu (jei gali trukdyti suvokti dokumentų esmę). Be tikslumo, administracinis stilius išsiskiria logiškumu (Bitinienė, 1998; Vladarskienė, 2004, 27), glaustumu, intelektualumu (Župerka, 1983, 106), nors tai ir nelaikoma specifiniais šio stiliaus požymiais.

Kita vertus, administraciniam stiliui visiškai nebūtinai individualumas, vaizdingumas (Župerka, 2012, 85), būdingiausios savybės apima nuasmeninimą, kalbos automatizavimą ir „sausumą“ (Pikčilingis, 1971a, 328). Laima Pečkuvienė taip pat mini tikslumą ir tikslingumą kaip administracinio stiliaus požymius (Pečkuvienė, 2013). Administracinio

stiliaus tikslumo išraiška – ilgi išvardijimai, įvairių išimčių nurodymas, išsamus kiekvienos norminamos / reglamentuojamos situacijos apibūdinimas (Vladarskienė, 2005). Vis dėlto, kaip mano, pvz., Vladarskienė, administraciniam stiliui tipiškesni trumpesni sakiniai, vengiama tuščiažodžiauti, t. y. tas pats dalykas verčiau pasakomas keliais trumpais sakiniais nei vienu ilgu (Vladarskienė, 2005). Atrodytų, kad tikslumas ir glaustumas vienas kitam prieštarauja, tačiau administraciniam stiliui priskirtiniams tekstams būdinga įvairovė, todėl galima daryti prielaidą, kad šios savybės tipiškos ne visiems be išimties šio stiliaus tekstams, ypač kadangi tikslumas ir glaustumas nelaukomi specifiniais šio stiliaus bruožais, nors tikslumas administracinėje kalboje ir laikomas svarbesniu nei glaustumas (Vladarskienė, 2005).

Anot Kniūkštos, administracinį ir mokslinį FS sieja jų raiškos tikslumas, logiškumas, objektyvumas, aiškumas bei vaizdingų ir emocinių žodžių vengimas (Kniūkšta, 2005). Administracinio ir mokslinio stilių panašumą aiškumo ir logiškumo atžvilgiu taip pat mini ir Vladarskienė (2007). Abiejų stilių tekstuose gausu veiksmažodinių daiktavardžių ir konstrukcijų su jais (Vladarskienė, 2007). Kita vertus, administracinį stilių nuo mokslinio skiria jo standartiškumas / šabloniškumas (Žilinskienė, 2002b; Kniūkšta, 2005). Kitaip sakant, administraciniam stiliui būdinga standartiškos reikšmės (šabloninių) žodžių ir žodžių junginių, kuriuos Vladarskienė (2007, 58) vadina „kanceliarizmais“, gausa.

Vladarskienė, rašydama apie kalbos vartojimo klaidas administraciniuose tekstuose, teigia, kad lietuvių administracinei kalbai būdingas neutralumas ir oficialumas yra perimti iš rusų kalbos, kadangi ilgai dokumentai buvo rengiami pagal sovietinių dokumentų pavyzdžius, o mandagumo raiška tuo metu buvo laikoma „buržuazinėmis atgyvenomis“ (Vladarskienė, 2006, 48). Kita vertus, kai kurie šabloniniai junginiai, pavyzdžiui, prašymui pritarti, lietuvių administracinėje kalboje vartojami jau nuo pirmųjų lietuviškų kanceliarinių raštų pasirodymo XVI–XVII a. (Župerka, 2012, 85). Tad administracinės kalbos raidai įtakos turėjo įvairūs veiksniai – istoriniai, socialiniai, kultūriniai, ekonominiai.

Nors administraciniame stiliuje kalbos pagrindą sudaro bendrinės kalbos normos, čia kalba pritaikoma pagal vartojimo srities poreikius ir atliekamas funkcijas, tad žodžiai ir jų junginiai gali įgyti specifinių reikšmių (Vladarskienė, 2007). Be to, jei kalbos reiškinys (pavyzdžiui, junginys ar konstrukcija) paplinta administracinėje kalboje, jo vartoseną turi polinkį plėstis. Tokiu būdu administracinei kalbai būdingi raiškos ypatumai ima reikštis ir kituose stiliuose, ypač publicistiniame (Vladarskienė, 2007). Administracinės kalbos plitimas yra susijęs ir su jau nekart minėtu jos

standartiškumu / šabloniškumu, kuris paprastai vertinamas teigiamai, nes palengvina dokumentų tvarkymą.



1 pav. Administracinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu

Apibendrinant lietuvių kalbos administracinio FS svarbiausias savybes iš kokybinės perspektyvos, išryškėjo šios lietuvių mokslininkų išskirtos ypatybės (žr. 1 pav.): standartiškumas / šabloniškumas, objektyvumas, oficialumas, aiškumas, glaustumas, tikslingumas, konkretumas, tikslumas, logiškumas, neutralumas, intelektualumas. Šias savybes iš dalies galima susieti su šios disertacijos kiekybinėje lietuvių kalbos administracinio stiliaus analizėje (4.1 sk.) naudotais kiekybiniais indikatoriais (1.4 sk.). Kadangi FS tyrę mokslininkai santykiškai nedaug analizavo, kuo ir kaip kokybiniais metodais išskirtos savybės pasireiškia tekste (disertacijos kontekste minėtini kalbos dalių pasiskirstymo (Žilinskienė, 2002c), leksikos įvairovės (Žilinskienė, 2002b) bei sakinio ilgio ir tipų pasiskirstymo (Bitinienė, 2002; Vladarskienė, 2004) tyrimai), toliau pateikiamos administracinio stiliaus kokybinių ir kiekybinių charakteristikų galimos sąsajos yra disertacijos autorės interpretacija.

Standartiškumą / šabloniškumą galima susieti su turinio / teminių¹⁴ žodžių ir žodžių formų vartoseną – jų gausa ir pasikartojimais (žymi indikatoriai R_1 (plačiau žr. 1.4.2.2 sk.) ir RR_{mc} (plačiau žr. 1.4.2.3 sk.)) bei dažnumu į funkcinius žodžius ir žodžių formas panašių turinio / teminių žodžių ir žodžių formų vartojimo (žymi indikatorius TK (plačiau žr. 1.4.2.5 sk.)) ypatumais.

¹⁴ Turinio / teminiai žodžiai ir žodžių formos – tai pagrindinėms kalbos dalims priklausantys žodžiai ir žodžių formos. Pagrindinėmis kalbos dalimis laikomi daiktavardžiai, būdvardžiai, skaitvardžiai, veiksmažodžiai,rieveksmiai (Utkā, 2004, 36–37).

Oficialumas tekste paprastai pasireiškia įvardžių ir veiksmažodžių formomis (Šlepikienė ir Linkevičienė, 2014). Oficialumo raišką kiekybinės lingvistikos perspektyvoje galima susieti su indikatoriumi *a* (plačiau žr. 1.4.2.1 sk.): kuo šio indikatorius vertė mažesnė, tuo didesnę teksto dalį sudaro labai dažni žodžiai ir žodžių formos, kuriems priklauso ir įvardžiai. Glaustumo raišką galima įžvelgti per neišplėtotus gramatinius centrus tekste ir tai įvertinanti indikatorius VA (plačiau žr. 1.4.3.3 sk.).

Apskaičiuojant kiekybinių indikatorius vertes, gaunami skaitiniai įverčiai, kurių interpretacija kalbos atžvilgiu priklauso nuo tiriamos problemos, sociolingvistinių, pragmatinių ir kitų ypatumų, tiriamo kalbos lygmens (pavyzdžiai, leksikos, morfologijos, sintaksės), kalbos atmainos ir kt. Tad kai kokybinės funkcinio stiliaus charakteristikos interpretuojamos per kiekybinę stiliaus raišką, neretai tie patys indikatorius gali būti naudojami apibūdinti skirtingoms kokybinėms charakteristikoms. Pavyzdžiui, administracinio stiliaus intelektualumo ir konkretumo raišką galima iš dalies susieti su indikatoriais VŽFI (plačiau žr. 1.4.3.1 sk.), VA (plačiau žr. 1.4.3.3 sk.) ir *a* (plačiau žr. 1.4.2.1 sk.). Pirmasis indikatorius leidžia įvertinti tekstą pagal sudėtingumą (t. y. kaip lengva ar sunku tekstą suprasti), o administracinio stiliaus kontekste tai būtų, pavyzdžiui, terminų gausa, kurie neretai yra santykiškai ilgi žodžiai, dėl ko tekstas tampa sudėtingesnis. Antrasis indikatorius leidžia įvertinti teksto sintaksinį kompleksiskumą. Trečiasis indikatorius naudojamas labai dažnų (paprastai funkcinų) ir žodžių ir žodžių formų santykiui tekste nustatyti. Vienas galimų intelektualumo ir konkretumo raiškos būdų galėtų būti didesnis retesnių (teminių / turinio) žodžių ir žodžių formų santykis tekste lyginant su labai dažnais (labai dažni žodžiai ir žodžių formos paprastai būna funkciniai¹⁵).

Panašiai tikslumo kokybinę charakteristiką kiekybiškai iš dalies galima analizuoti per ilgesnius žodžius ir žodžių formas (dažnai būdinga terminams, faktinei informacijai) – žymi jau minėtas indikatorius VŽFI (plačiau žr. 1.4.3.1 sk.), turinio / teminių žodžių ir žodžių formų įvairovę (žymi indikatorius R_1 (plačiau žr. 1.4.2.2 sk.)), dažnumu funkciniais bemaž prilygstančių turinio / teminių žodžių ir žodžių formų gausumą (žymi indikatorius TK, plačiau žr. 1.4.2.5 sk.). Disertacijos autorė pastebėjo, kad administracinių tekstų pavadinimams būdingi būdvardžiai, kurie irgi gali būti siejami su tikslumo charakteristika. Būdvardžių gausumą tekste leidžia įvertinti indikatorius D (plačiau žr. 1.4.3.2 sk.). Vadovaujantis panašia

¹⁵ Funkciniais žodžiais laikomi prielinksniai, dalelytės, jungtukai (Utka, 2004, 36) bei įvardžiai (Utka, 2005, 51).

interpretacija, logiškumą galima iš dalies apibūdinti per didesnę labai dažnų žodžių ir žodžių formų (tarp jų patenka jungtukai, dalelytės,) santykį tekste lyginant su retesniais, paprastai – turinio / teminiais (žymi jau minėtas indikatorius a ; plačiau žr. 1.4.2.1 sk.). Objektyvumas sietinas su indikatoriumi TK (plačiau žr. 1.4.2.5 sk.) – turinio / teminių žodžių ir žodžių formų, savo dažnumu nedaug nusileidžiančių labai dažniems (paprastai – funkciniais) žodžiams ir žodžių formoms, kas gali būti interpretuojama, pavyzdžiui, gausiu faktinės, dalykinės informacijos pateikimu tekste.

Tikslingumas nurodo į tikslą, tad galima daryti prielaidą apie veiksmazodžių sąlygišką gausumą, susijusį su tikslo siekimu – veiksmu, ką iš dalies leidžia įvertinti indikatorius A (plačiau žr. 1.4.3.2 sk.). Panašiai neutralumas galėtų būti sietinas su indikatoriais A ir D (plačiau žr. 1.4.3.2 sk.) bei balansu tarp jų, kitaip – balansu tarp tekste esančių veiksmazodžių ir būdvardžių. O aiškumas – nepagrindinė administracinio stiliaus savybė – galėtų iš dalies būti įvertinamas per labai dažnų žodžių ir žodžių formų (jungtukų, dalelyčių) vartoseną (labai dažnų ir retesnių žodžių ir žodžių formų santykį tekste leidžia įvertinti indikatorius a (plačiau žr. 1.4.2.1 sk.), turinio / teminių žodžių ir žodžių formų vartojimo ypatumus, tokius kaip įvairovė ir pasikartojimų gausa (žymi indikatoriai R_1 (plačiau žr. 1.4.2.2 sk.) ir RR_{mc} (plačiau žr. 1.4.2.3 sk.)). Disertacijoje pristatomame tyrime FS kalbinė raiška tirama per kiekybinius indikatorius, o kiekybinių ir kokybinių požymių sąsajos, kad ir dalinės, palengvina tyrimo metu gautų rezultatų interpretavimą.

1.2.3 Lietuvių kalbos mokslinio stiliaus ypatybės

Kaip jau minėta, tikėtina, kad panašūs požymiai būdingi ir kitų kalbų moksliniams stiliams, tačiau disertacijos tikslas, kaip jau minėta, yra iširti lietuvių kalbos FS, tad šiame poskyryje apžvelgiami tik lietuvių mokslininkų darbai. Mokslinis stilius plėtojamas mokslinės veiklos srityje. Mokslinio stiliaus ypatybes ir dominuojančias kalbos priemones lemia tiek vartojimo sritis, tiek turinys bei kalbos funkcijos. Kitaip sakant, mokslinis stilius vartojamas naujų žinių pateikimui, jų kaupimui ir sisteminimui bei perdavimui, t. y. pagrindinė yra pranešimo / informacinė funkcija (Župerka, 2012, 81). Kita vertus, pagal Utkos kiekybinio tekstų funkcijų nustatymo tyrimo duomenis, moksliniam stiliui (arba, kaip teigia Utko, akademinės prozos superžanrui) būdingiausia deskriptyvumo funkcija, pasireiškianti ilgais sakiniais, beasmenėmis konstrukcijomis ir sąlygiškai turtingesniu žodynu (Utko, 2004, 107). Moksliniam stiliui būdingos specifinės stilistinės savybės – tai abstraktumas, minčių dėstymo nuoseklumas, dalykinis

tikslumas, objektyvumas, glaustumas (Bitinienė, 1983, 9), patikimumas, aiškumas (Petrėnienė, 2003). Anot Bitinienės, mokslinio stiliaus tekstams būdingas logiškumas, intelektualumas, samprotavimas ir įrodinėjimas, nes juose pateikiami žmogaus intelektinės veiklos rezultatai (Bitinienė, 2000). Bitinienės teigimu, ilgi sudėtingi sakiniai mokslinio stiliaus tekstuose paprastai rodo samprotavimo pagrindą, o trumpi ir ilgi vientisiniai sakiniai rodo pranešimo (informacinės) ir ekspresyvinės funkcijų sąryšį (Bitinienė, 2000).

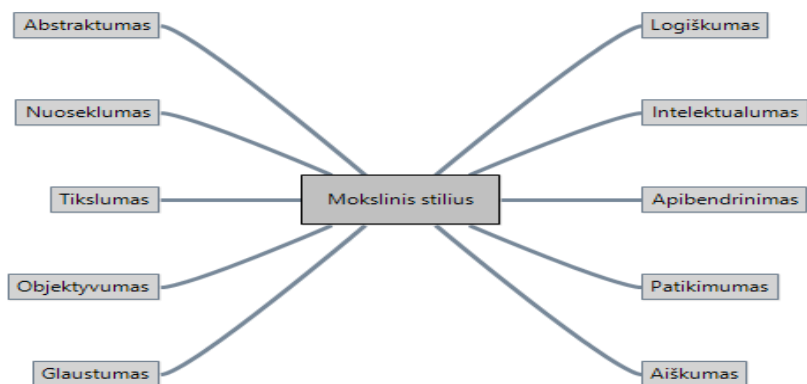
Mokslinio stiliaus tekstuose leksinių priemonių dažnumai bei vartosenos ypatumai gali būti siejami su šiam stiliui būdingais jau minėtais nekalbiniais požymiais – abstraktumu (apibendrinimu), logiškumu, nuoseklumu (Bitinienė, 2009). Vientisiniai sakiniai moksliniame stiliuje gali būti interpretuojami leksinių pasikartojimų atžvilgiu: jie sukuria žodžių perteklių, naudingą perduodant žinias mokomojo postilio tekstuose. Bitinienė tai įvardija kaip „sutirštintą rišlumą“ (Bitinienė, 2009, 24). Kalbat apie mokslinį stilių, pasikartojimais laikomi ne tiek paprasti žodžių pasikartojimai, bet greičiau tos pačios informacijos perteikimas įvairiais būdais (Bitinienė, 2009). Župerkos akcentuojama, kad moksliniam stiliui nebūdingas vaizdingumas (Župerka, 1983, 102; Župerka, 2012, 81), tačiau šios savybės raiška moksliniame stiliuje priklauso nuo mokslo šakos bei postilio, tad tai nėra viena pagrindinių šio stiliaus charakteristikų. Pavyzdžiui, moksliniuose tekstuose galima rasti vaizdingų ir metaforinių terminų (Vladarskienė, 2012). Be to, nors mokslo populiarinamųjų tekstų, taip pat priklausančių moksliniam stiliui, pagrindinė funkcija – pranešimo, tačiau, kad tai būtų daroma patraukliai ir įtraukiančiai, kartu realizuojamos ir ekspresinė bei apeliacinė funkcijos (Župerka, 2012, 83). Tad kartu su dalykiniu tikslumu, logiškumu ir faktų patikimumu mokslo populiarinamiesiems tekstams būdingi vaizdingumas, emocionalumas ir subjektyvus vertinimas (Petrėnienė, 2003).

Lietuvių kalbos stilistikos darbuose neretai laikomasi tik vienos krypties mokslinio stiliaus tipologijos ir atsiskoma (arba teikiama menkesnė svarba) mokslinio stiliaus postilių, paremtų konkrečiau mokslo specifika (Pikčilingis, 1971a; Župerka, 2012, 80). Tad lietuvių kalbos stilistikos darbuose skiriami trys mokslinio stiliaus postiliai: teorinis / specialusis (mokslinis siaurąją prasme; apima, pavyzdžiui, tezes, straipsnius, monografijas ir pan.) ir mokslo populiarinimo (Pikčilingis, 1971a, 315) bei mokomasis (pamokos, paskaitos, pranešimai) (Župerka, 1983, 102). O, pavyzdžiui, Ilja Galperin darbe apie anglų kalbos stilistiką kaip mokslinio stiliaus postilius išskiria tikslųjų mokslų postilį, humanitarinių mokslų postilį ir mokslo populiarinamąjį postilį (Galperin, 1977, 34). Kitaip sakant, Galperino mokslinio stiliaus postilių tipologija labiau remiasi kalbinių požymių skirtumais, nes informacijos

pateikimo struktūra, žodynas bei vartojamos sintaksinės konstrukcijos skirtingoms mokslo kryptims reikšmingai skiriasi.

Galima sakyti, jog Galperin, apibrėždamas anglų kalbos FS ir postilius, yra linkęs pabrėžti kalbinius veiksnius (stilistines ypatybes ir kalbos priemones), o ne nekalbinius – vartojimo sritį ir funkciją ar funkcijas. Kadangi jau pačiame FS pavadinime yra žodis „funkcinis“, akcentuojant kalbinius veiksnius tolstama nuo tekstų kategorizavimo pagal jų atliekamas funkcijas ir artėjama prie žanrinės tipologijos. Lietuvių kalbos mokslinio stiliaus tekstų skirtumams pagal mokslo kryptis apibrėžti laikytasi kitokios strategijos – šie skirtumai kokybinėmis įžvalgomis apibrėžiami per turinį / tematiką. Tad teoriniam / specialiajam postiliui priklausantys tekstai dar šiuo atžvilgiu skirstomi į humanitarinių mokslų, gamtos mokslų ir technikos bei matematikos mokslų (Župerka, 2012, 80), taip išlaikant funkcijos dominavimą apibrėžiant FS, tačiau pripažįstant kalbinės raiškos skirtumų egzistavimą. Sudarant tikslinį mokslinio funkcinio stiliaus tekstyną MST atsižvelgta į kalbinės raiškos skirtumų egzistavimą skirtingų mokslo krypčių tekstuose, tad tekstyną sudarantys tekstai apima ne tik visus lietuvių kalbos mokslinio stiliaus postilius (teorinį / specialųjį, mokslo populiarinamąjį ir mokomąjį), bet ir visas mokslo kryptis (plačiau žr. 2.3.2 sk.).

Kalbant apie mokslinio ir kitų šioje disertacijoje tiriamų lietuvių kalbos FS (administracinio ir publicistinio) santykį, mokslinio stiliaus tekstams būdingesni trumpi sakiniai, o tai, anot Petrėnienės (2011, 158), rodytų šių tekstų artėjimą prie publicistinio stiliaus, ypač mokslo populiarinamojo postilio kontekste. Mokslinis ir administracinis funkciniai stiliai turi panašumų, tačiau juos skiria šabloniškumo / standartiškumo lygmuo, kuris administraciniuose tekstuose yra daug labiau išreikštas nei moksliniuose tekstuose (Žilinskienė, 2002a; Kniūkšta, 2005). Kitaip sakant, nors šabloniški žodžių junginiai, frazės, morfologinės ir sintaksinės struktūros yra tipiškesnės administraciniams tekstams, moksliniams jie taip pat būdingi, tik sustabarėjimo laipsnis yra daug mažesnis.



2 pav. Mokslinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu

Apibendrinant lietuvių kalbos mokslinio FS svarbiausias savybes iš kokybinės perspektyvos, lietuvių mokslininkai išskyrė šias ypatybes (žr. 2 pav.): abstraktumas, logiškumas, nuoseklumas, intelektualumas, tikslumas, apibendrinimas, objektyvumas, patikimumas, glaustumas, aiškumas. Šias savybes atsargiai ir iš dalies galima susieti su kiekybinėje lietuvių kalbos mokslinio stiliaus analizėje (4.2 sk.) naudotais kiekybiniais indikatoriais (1.4 sk.). Kadangi, kaip minėta, FS tyrę mokslininkai santykiškai mažai analizavo, kuo ir kaip kokybiniais metodais išskirtos savybės pasireiškia tekste kiekybiškai (disertacijos kontekste paminėtini sakinių ilgių ir tipų (Bitinienė, 2000; 2007; 2009; Petrėnienė, 2011), kalbos dalių pasiskirstymo (Žilinskienė, 2002a; Bitinienė, 2007b; 2009; Petrėnienė, 2011) tyrimai), toliau pateikiamos mokslinio stiliaus kokybinių ir kiekybinių charakteristikų galimos sąsajos yra disertacijos autorės interpretacija. Primintina, kad apskaičiuojant kiekybinių indikatorių, naudotų šioje disertacijoje pristatomame tyrime, vertes, gaunami skaitiniai įverčiai, kurių interpretacija kalbos atžvilgiu priklauso įvairių faktorių (tiriamos problemos, sociolingvistinių, pragmatinių ir kitų ypatumų, tiriamo kalbos lygmens, kalbos atmainos ir pan.), todėl neretai tie patys indikatoriai gali būti naudojami apibūdinti skirtingoms kokybinėms charakteristikoms.

Abstraktumą ir apibendrinimą galima sieti su abstrakčios reikšmės žodžiais, kurie perteikia apibendrintą informaciją (Bitinienė, 2009)), jų gausa. Ieškant sąsajų su disertacijoje pristatomame tyrime naudotais kiekybiniais indikatoriais randame, kad rodomųjų įvardžių, kurie, anot Bitinienės (2009), kartu su apibendrintos reikšmės žodžiais išreiškia informacijos apibendrinimą, gausumą būtų galima įvertinti per indikatorių *a* (plačiau žr. 1.4.2.1 sk.), kuris vertina labai dažnų ir retesnių žodžių ir žodžių formų santykį tekste (įvardžiai patenka tarp labai dažnų žodžių ir žodžių formų). O

apibendrintos reikšmės žodžių gausumą ir dažnumą turėtų leisti įvertinti indikatorius TK (plačiau žr. 1.4.2.5 sk.). Vadovaujantis panašia interpretacija, logiškumą ir nuoseklumą galima iš dalies įvertinti per labai dažnų žodžių ir žodžių formų proporciją tekste lyginant su retesniais, paprastai – turinio / teminiais žodžiais ir žodžių formomis¹⁶ (žymi jau minėtas indikatorius a ; plačiau žr. 1.4.2.1 sk.). Didesnis kiekis labai dažnų žodžių ir žodžių formų (paprastai – funkcinių¹⁷, tad tarp jų patenka jungtukai, dalelytės) tekste, lyginant su retesniais, galėtų rodyti didesnę logiškumo ir nuoseklumo raišką. Analogišku principu tikslumo kokybinę charakteristiką kiekybiškai iš dalies galima tirti per žodžių ir žodžių formų ilgį (naudojamas jau minėtas indikatorius VŽFI (plačiau žr. 1.4.3.1 sk.), laikantis prielaidos, kad ilgesni žodžiai ir žodžių formos dažnai būdingi terminams, faktinei informacijai. Be to, tikslumo raišką būtų galima iš dalies įvertinti per turinio / teminių žodžių ir žodžių formų įvairovę (žymi indikatorius R_1 (plačiau žr. 1.4.2.2 sk.)), dažnumu funkciniais bemaž prilygstančių turinio / teminių žodžių ir žodžių formų gausumą (žymi indikatorius TK, plačiau žr. 1.4.2.5 sk.).

Mokslinio stiliaus intelektualumo raišką galima atsargiai sieti su indikatoriais VŽFI (plačiau žr. 1.4.3.1 sk.), VA (plačiau žr. 1.4.3.3 sk.) ir a (plačiau žr. 1.4.2.1 sk.). Šio stiliaus kontekste intelektualumo požymis galėtų būti, pavyzdžiui, terminų gausa. Daroma prielaida, kad terminai neretai yra santykiškai ilgi žodžiai, dėl ko tekstas tampa sudėtingesnis (suprantamumo prasme), o tai leidžia įvertinti VŽFI. Kitas intelektualumo požymis galėtų būti teksto sintaksinis kompleksiskumas, kurį leidžia įvertinti indikatorius VA. Intelektualumo raišką iš dalies būtų galima susieti ir su didesniu, lyginant su labai dažniais (labai dažni žodžiai ir žodžių formos paprastai būna funkciniai), retesnių (teminių / turinio) žodžių ir žodžių formų santykiu tekste. Objektyvumą galima susieti su indikatoriumi TK (plačiau žr. 1.4.2.5 sk.), skirtą įvertinti turinio / teminių žodžių ir žodžių formų, savo dažnumu nedaug nusileidžiančių labai dažniems (paprastai – funkciniais) žodžiams ir žodžių formoms, kiekį tekste. Objektyvumo kontekste šių žodžių ir žodžių formų gausa galėtų rodyti faktinės ir / ar dalykinės informacijos gausą tekste.

Patikimumo raišką moksliniame stiliuje sustiprina terminai, faktinė informacija, kuriems, daroma prielaida, būdingesni ilgesni žodžiai ir žodžių formos. Tad patikimumui įvertinti gali būti naudojamas jau minėtas indikatorius VŽFI (plačiau žr. 1.4.3.1 sk.). Faktinės informacijos raišką taip pat galima išvelgti teminių žodžių ir žodžių formų įvairovėje (žymi

¹⁶ žr. 14-ą išnašą.

¹⁷ žr. 15-ą išnašą.

indikatorius R_1 (plačiau žr. 1.4.2.2 sk.)) bei dažnumu funkciniais bemaž prilygstančių turinio / teminių žodžių ir žodžių formų vartosenoje (žymi indikatorius TK, plačiau žr. 1.4.2.5 sk.). Glaustumo raišką galima būtų įžvelgti per neišplėtotus gramatinius centrus tekste, ką leidžia įvertinti indikatorius VA (plačiau žr. 1.4.3.3 sk.). O aiškumą iš dalies galima būtų sieti su labai dažnų žodžių ir žodžių formų (jungtukų, dalelių) vartoseną ir jų santykiu su retesniais, t. y. turinio / teminiais (tai leidžia įvertinti indikatorius a (plačiau žr. 1.4.2.1 sk.)) bei turinio / teminių žodžių ir žodžių formų vartojimo ypatumus – įvairovę ir pasikartojimų gausą (tai žymi indikatoriai R_1 (plačiau žr. 1.4.2.2 sk.) ir RR_{mc} (plačiau žr. 1.4.2.3 sk.)). Nors disertacijoje pristatomame tyrime FS kalbinė raiška tiriama per kiekybinius indikatorius, tačiau kiekybinių ir kokybinių požymių sąsajos palengvina tyrimo rezultatų interpretavimą.

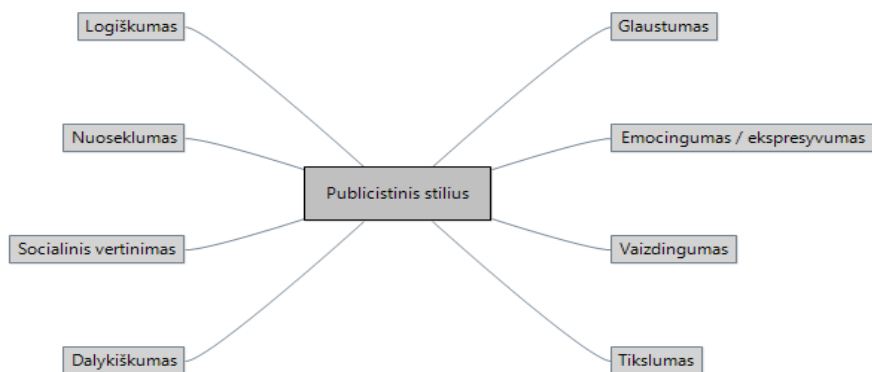
1.2.4 Lietuvių kalbos publicistinio stiliaus ypatybės

Tikėtina, kad panašūs požymiai būdingi ir kitų kalbų publicistiniams stiliams, tačiau kadangi disertacijoje, kaip jau minėta, siekiama ištirti lietuvių kalbos FS, tad šiame poskyryje apžvelgiami tik lietuvių mokslininkų darbai. Tad publicistinis FS yra dar vienas šioje disertacijoje aprašomo tyrimo objektas. Šio stiliaus specifika – informuoti ir įtikinti (daryti poveikį), pateikti socialinį vertinimą, realizuojama ir emocinio poveikio funkcija. Mokslinio stiliaus elementai, randami publicistiniame stiliuje, apima loginę naujienų pateikimo struktūrą, teksto nuoseklumą. Publicistiniame stiliuje siekiama vartoti skaitytojui kuo patrauklesnę ir įdomesnę, kartu ir kuo informatyvesnę kalbą (Bitinienė, 2007c). Šiam stiliui būdingas mokslinio stiliaus dalykiškumas, glaustumas, nuoseklumas, logiškumas, tačiau informacija dažnai pateikiama panašia į meninio stiliaus forma (Smetonienė, 2013). Remiantis A. Utkos kiekybinio tekstų funkcijų nustatymo tyrimo rezultatais, publicistiniam stiliui (Utkos tyrimo medžiagoje ši stilių atitiktų spaudos superžanras) būdingiausia tempatyvumo funkcija, kurios raišką sudaro laiką žyminčių žodžių gausa (Utko, 2004, 107—108). Ieškant sąsajų su kokybiniais publicistinio stiliaus požymiais, tempatyvumo raiška galėtų sietis su logine naujienų pateikimo struktūra.

Anot Župerkos, svarbiausia publicistinio stiliaus vartojimo sritis yra periodinė spauda (Župerka, 2012, 86), tad šio stiliaus tekstai yra skirti plačiajai visuomenei, kai norima aptarti įvairius svarbius klausimus – politinius, socialinius ar kitus. Pikčilingis išskiria publicistinio stiliaus įtikinimo funkciją (Pikčilingis, 1971a, 307), o Župerka – pranešimo ir poveikio funkcijas (Župerka, 2012, 86), nes publicistiniuose tekstuose siekiama ne tik pateikti informaciją, bet ir formuoti visuomenės nuomonę. Be to, publicistiniam stiliui būdingas socialinis vertinimas (Bitinienė, 2007c, 16; Župerka, 2012, 86—87), nes šio stiliaus tekstų turinį sudaro aktualios visuomenei problemos, apie kurias visuomenės nariai yra linkę išreikšti savo nuomonę. Kadangi publicistiniam stiliui būdingi tiek meninio (emocingumas, vaizdingumas, ekspresyvumas), tiek mokslinio (dalykiškumas, logiškumas, tikslumas, aiškumas, glaustumas) funkcinių stilių ypatumai, tad ir kalbinės priemonės apima tiek logines, tiek emocines (Župerka, 2012, 87). Pavyzdžiui, okazionaliųjų posakių ir frazeologizmų vartojimas publicistiniame stiliuje atskleidžia frazeologijos dinamiškumą, prisitaikymo prie kasdienių įvykių galimybes, t. y. kuo išraiškingiau pateikiama informacija, tuo tekstas labiau atkreipia skaitytojo dėmesį (Butkutė, 2010, 153), o demonstratyvinė rodomoji referencija, reiškiamą įvardžiais *šis*, *šitas*, *tas* (Česnulienė, 2012), prisideda

prie logiškumo, nuoseklumo raiškos. Paminėtina, kad publicistinio stiliaus teksto autorius neretai išsako ne savo asmeninę nuomonę, o apibendrina daugelio visuomenės narių nuomones ir požiūrius (Stukaitė, 2005, 5).

Priešingai nei kiti lietuvių kalbos FS, Publicistinis stilius nėra skirstomas į apibrėžtus postilius, daugiau remiamasi žanrine tipologija. Pikčilingis pastebi retorinio postilio ir meninio (beletristinio) FS panašumus (Pikčilingis, 1971a, 310), Župerka išskiria 3 publicistinio stiliaus postilius pagal pranešimo ir poveikio funkcijų santykį: informacinį (vyrauja pranešimo funkcija; apima, pavyzdžiui, žinutę, ataskaitą, reportažą), analitinį (būdingas pranešimo ir poveikio funkcijų balansas; apima, pavyzdžiui, straipsnį, apybraižą) ir ekspresyvųjį (vyrauja poveikio funkcija, kuri aptinkama, pavyzdžiui, miniatiūrose, recenzijose) (Župerka, 2012, 87). Informacinis postilis išsiskiria daiktavardžių gausa, ekspresyvusis – dideliu kiekiu veiksmazodžių ir jungtukų, o analitiniame postilyje dažnesni įvardžiai ir būdvardžiai (Stukaitė, 2005, 90). I. Smetonienė išskiria 4 publicistinio stiliaus postilius: informacinį, analitinį, ekspresyvųjį ir dialoginį (Smetonienė, 2013c). Pirmieji 3 postiliai sutampa su Župerkos publicistinio stiliaus postiliais, ketvirtasis apima dialoginius žanrus, pavyzdžiui, interviu, pokalbius. Galima teigti, kad šnekamoji kalba yra skirta masinei komunikacijai kasdienėse veiklose, tad ji didina informacijos prieinamumą, o dialogizavimas publicistiniuose tekstuose juos kaip tik ir priartina prie šnekamosios kalbos (Bitinienė, 2004). Župerkos ekspresyvųjį publicistinio stiliaus postilį A. Bitinienė vadina meniniu (Bitinienė, 2003).



3 pav. Publicistinio stiliaus požymiai – kokybinių tyrimų duomenų pagrindu

Apibendrinant lietuvių kalbos publicistinio FS svarbiausias savybes iš kokybinės perspektyvos, lietuvių mokslininkai išskyrė šias ypatybes (žr. 3 pav.): logiškumas, glaustumas, nuoseklumas, emocingumas / ekspresyvumas,

socialinis vertinimas, vaizdingumas, dalykiškumas, tikslumas. Šias savybes galima, bent iš dalies, susieti su kiekybinėje lietuvių kalbos publicistinio stiliaus analizėje (4.3 sk.) naudotais kiekybiniais indikatoriais (1.4 sk.). Kaip jau buvo minėta, lietuvių kalbos FS tyrę mokslininkai santykiškai mažai analizavo, kaip kokybiniais metodais išskirtas savybes tirti kiekybiškai (disertacijos kontekste paminėtini sakinių ilgių ir tipų (Bitinienė, 2004; 2008; Leonavičienė, 2005; 2010), leksikos ir kalbos dalių pasiskirstymo (Čičirkaitė, 2005; Žilinskienė, 2001; 2002c) tyrimai), toliau pateikiamos galimos sąsajos tarp publicistinio stiliaus kokybinių ir kiekybinių charakteristikų yra disertacijos autorės interpretacija. Primintina, kad kiekybinių indikatorių, naudotų šioje disertacijoje pristatomame tyrime, skaitiniai įverčiai gali būti interpretuojami įvairiai, nes interpretacija kalbos atžvilgiu priklauso įvairių faktorių, pavyzdžiui, tiriamos problemos, tiriamo kalbos lygmens, kalbos atmainos, sociolingvistinių, pragmatinių ir kitų ypatumų. Tad neretai ta pati indikatorių kombinacija gali būti naudojama apibūdinti skirtingoms kokybinėms charakteristikoms.

Publicistiniame stiliui būdingus logiškumą ir nuoseklumą galima iš dalies įvertinti per labai dažnų ir retesnių žodžių ir žodžių formų santykį tekste (labai dažni žodžiai ir žodžių formos paprastai būna funkciniai, o retesni – turinio / teminiai (žymi jau minėtas indikatorius a ; plačiau žr. 1.4.2.1 sk.)). Didesnis, lyginant su retesniais, kiekis labai dažnų žodžių ir žodžių formų, tarp kurių patenka ir jungtukai, dalelytės, tekste galėtų rodyti didesnę logiškumo ir nuoseklumo raišką. Glaustumo apraiškas tekste būtų galima išvelgti per neišplėtotus gramatinius centrus tekste, o tai leidžia įvertinti indikatorius VA (plačiau žr. 1.4.3.3 sk.).

Panašiai dalykiškumo ir tikslumo charakteristiką, laikantis prielaidos, kad ilgesni žodžiai ir žodžių formos būdingesni faktinei informacijai bei terminams, kiekybiškai iš dalies galima sieti su tekstus pagal jų sudėtingumą (žodžių ilgių atžvilgiu, t. y. tekstai, kuriuose dominuoja trumpesni žodžiai ir žodžių formos laikomi paprastesniais, o kuriuose vyrauja ilgesnis – sudėtingesniais) leidžiančiu įvertinti indikatoriumi VŽFI (plačiau žr. 1.4.3.1 sk.). Tikslumo raišką dar galima analizuoti per turinio / teminių žodžių ir žodžių formų įvairovę (žymi indikatorius R_1 (plačiau žr. 1.4.2.2 sk.)) ir dažnumu funkciniais beveik prilygstančių turinio / teminių žodžių ir žodžių formų vartoseną (naudojamas indikatorius TK, plačiau žr. 1.4.2.5 sk.).

Vienos esminių publicistinio stiliaus savybių – socialinio vertinimo – raišką kiekybiškai iš dalies būtų galima sieti su vertinamųjų būdvardžių vartoseną, kurią leistų užfiksuoti indikatorius D (plačiau žr. 1.4.3.2 sk.). O emocingumo / ekspresyvumo ir vaizdingumo raišką galima būtų susieti su indikatorių D ir MATTR kombinacija (MATTR yra modifikuotas iteracijos

indeksas, plačiau žr. 1.4.2.4.sk.), t. y. šioje kombinacijoje D leistų įvertinti teksto deskriptyvumą (kadangi ši charakteristika yra priešinga aktyvumui, išreiškiamam per veiksmožodžių vartoseną, deskriptyvumas sietinas ir su ekspresyvumu, vaizdingumu, nors, žinoma, šios sąsajos yra dalinės) per būdvardžių ir veiksmožodžių santykį, o MATTR – įvertinti skirtingų ir visų būdvardžių santykį, kitaip – deskriptyvumo „srautą“ tekste.

Disertacijoje pristatomame tyrime FS kalbinė raiška tirama per kiekybinius indikatorius, tačiau kiekybinių požymių siejimas su kokybiniais palengvina rezultatų, gautų taikant kiekybinius metodus, interpretavimą. Tad, kadangi tyrimas remiasi kiekybinės lingvistikos metodais, toliau pristatoma trumpa kiekybinės lingvistikos istorija, disertacijoje pristatomame tyrime naudoti kiekybiniai indikatoriai bei trumpa tyrimų, kuriuose naudoti kiekybiniai indikatoriai, apžvalga.

1.3 Kiekybinė lingvistika: trumpa istorija

Šioje disertacijoje aprašomas tyrimas atliekamas taikant kiekybinės lingvistikos (toliau – KL) metodus, tad šiame skyriuje trumpai pristatoma kiekybinės lingvistikos istorija. Ruina Chen ir Haitao Liu (2014) išskiria 3 KL kaip disciplinos raidos etapus: 1) kiekybinės lingvistikos gimimą kartu su Džordžo Kingslio Zipf'o publikacija „The Psycho-Biology of Language – An Introduction to Dynamic Philology“ 1935 metais ir šio tyrėjo išpopuliarintą Zipf'o dėsnį, 2) plėtros etapą, kurio svarbiausias tyrėjas yra Gabriel Altmann (nuo 1970 m. iki dabar) 3) brandos etapą, kurį reprezentuoja Reinhard Köhler (nuo 1993 m. iki dabar), taikantis sinergetikos¹⁸ prieigas kalbos tyrimams, kartu integruodamas atskirus kalbos. Ši kategorizacija neapima ankstyvojo KL laikotarpio iki minėto Zipf'o straipsnio. Be to, paminėtina, kad tarp KL istorinių etapų, iš esmės apibrėžiamų per taikomus metodus, nėra aiškių ribų. Kitaip tariant, atsiradus sudėtingesniems metodams, paprastesniųjų taikymas nenunyksta, jie ir toliau taikomi, tik kartais mažesniu mastu ir / arba kartu su naujaisiais metodais. Tolesniuose poskyriuose pateikiama trumpa KL istorija nuo XIX a. pab. iki šių dienų.

¹⁸ Sinergetika – mokslas apie sistemų saviorganizacijos mechanizmus. Tad sinergetinė lingvistika – mokslas apie kalbos kaip sistemos bei šią sistemą sudarančių struktūrų „saviorganizacijos mechanizmus“ (Köhler, 1993, 41).

1.3.2 Ankstyvasis kiekybinės lingvistikos laikotarpis (iki 1935 m.)

KL atsiradimo pradžia – XIX a. pab. – XX a. pirmą pusę. Pirmieji tyrimai, kuriuose buvo skaičiuojami tiriami kalbos ar tekstų vienetai, publikuoti XIX a. Tuo metu Vokietijoje tokius tyrimus vykdė Ernst Förstemann (Förstemann, 1852; Best, 2006), Rusijoje – Viktoras Bujankovskis, Prancūzijoje – Benjamin Bourdon, JAV – Lucius Adelno Sherman (Köhler, 2005a, 3). Šie autoriai tyrė kalbos vienetų pasikartojimų dažnumus, siekdami, jais remdamiesi, aprašyti minėtus kalbos vienetus. Dėl disertacijos apimties ribojimų, išsamiau paminėti tik keli svarbiausi šio laikotarpio tyrėjai.

Vienas paminėtinų ankstyvojo KL laikotarpio tyrėjų yra Ernst Wilhelm Förstemann (1822–1906). Jis laikomas vienu statistikos taikymo kalbos tyrimuose pradininkų. Förstemann XIX a. vid. – XIX a. pab. laikotarpiu atliko įvairių lingvistinių fenomenų (garsų, gramatinių reiškinių, vardų) statistinius tyrimus (Best, 2006, 77–86; Schneider, 2006). Pirmasis, peržengęs aprašomosios analizės ribas, buvo rusų matematikas Andrejus A. Markovas, 1913 m. padėjęs pagrindus Markovo grandinių teorijai. Jo sukurtas matematinis modelis apibūdina sintagminę priklausomybę tarp tiriamų vienetų, tarp jų esant tiesiniam sąryšiui. Tirdamas Aleksandro Puškino eilėraščius Markovas pastebėjo, kad balsių ir priebalsių santykis žodyje kinta pagal raidės vietą jame, t. y. žodžio pradžioje ar pabaigoje (Basharin, Langville ir Naumov, 2004). Be to, šis santykis priklauso nuo to, ar prieš tai buvusi raidė buvo balsė ar priebalsė. Šie pasikartojimai leido sukurti Markovo grandinių ir paslėptuosius Markovo modelius, sėkmingai naudojamus automatiniam kalbos dalių atpažinimui tekste ir šiandien (Ramage, 2007; Martinez, 2012).

Kitas garsus ankstyvojo KL periodo mokslininkas buvo Otto Behaghel (1854–1936). Jis 1923–1932 m. suformulavo Behaghel'io dėsnius (Best, 2007, 80–86), apibūdinančius bazinius žodžių ir frazių pozicijų sakinyje principus, kurie tebėra svarbūs ir dabartiniuose temos ir remos tyrimuose (Collinge, 1985, 241). Vėliau kiekybinės lingvistikos plėtotę lėmė praktinės priežastys, tokios kaip pastangos pagerinti svetimosios kalbos mokymą ar stenografavimo sistemų optimizavimą (Köhler, 2005a, 3). Ankstyvieji žodyno tyrimo matematiniai modeliai publikuoti Jean-Baptiste Estoup (1912), George Udny Yule (1925) ir Edward Uhler Condon (1928) darbuose.

1.3.3 Nuo Zipf'o iki Altmann'o (nuo 1935 m. iki 1970 m.)

Pirmasis tyrėjas, kuris remdamasis teoriniais matematiniais modeliais bandė gauti kiekybinius rezultatus, buvo Džordžas Kingslis Zipf'as (George Kingsley Zipf) (Grzybek ir Köhler, 2011, vi; Köhler, 2012, 13). Jis analizavo kalbą ir tekstą taikydamas statistinius metodus bei suformulavo ir išgarsino Zipf'o dėsnį, apibrėžiantį ryšį tarp žodžių pasikartojimų dažnumų ir jų vietos dažniniame sąrašė. Šis dėsnis apibrėžia matematinę priklausomybę tarp žodžio dažnio ir jo vietos dažniniame sąrašė, taip pat tarp žodžio dažnio ir žodžių, turinčių tą dažnį, skaičiaus. Matematinė Zipf'o dėsnio formulė yra ši (Sicilia-Garcia, Ming ir Smith, 2003):

$$f = \frac{k}{r} \quad (1)$$

kur f žymi žodžio dažnį tekстыne, r – šio žodžio rangą, t. y. poziciją dažniniame sąrašė, o k žymi konstantą (pastovi skaitinė išraiška, nustatoma kiekvienam tekstynui).

Teigiama, kad šis dėsnis tinka bet kurios kalbos dažniniam sąrašėui, kitaip sakant – yra universalus. Šis dėsnis nepaneigtas, juo remiamasi ir šiandien. Plačiau apie Zipf'o dėsnį žr. 1.4.1 sk.

Svarbų indėlį į KL įnešė Prahos mokyklos tyrėjai: Vilém Mathesius (1882–1945), Bohimil Trnka (1895–1984), Josef Vachek (1909–1996) ir kt. (Uhlířová, 2005, 130). Šiuos tyrėjus domino fonologinės ir leksinės kalbos savybės anglų ir čekų bei kitose kalbose (Těšitelová, 1992, 15). Žinoma, tuo metu dar nebuvo technologinių pajėgumų, kurie leistų automatiškai apdoroti didžiulius duomenų kiekius, tačiau bandymai rasti įvairių kalbos fenomenų pasikartojimus ir dėsningumus, prisidėjo dedant pagrindus KL kaip disciplinos vystymuisi.

Greta Prahos mokyklos tyrėjų paminėtinas ir Sergejus Grigorjevičius Čebanovas (1897–1966) (Grzybek, 2007; Dinu ir Dinu, 2009) buvo vienas pirmųjų, bandžusių matematiškai suformuluoti dėsnį lingvistinio tyrimo rezultatams, kitaip sakant – apibūdinti pastovų ryšį tarp kalbos reiškinių. Laikoma, kad jis pirmasis, t. y. 1947 m., pasiūlė žodžių ilgių pasiskirstymo tekste modelį. Tuo pat metu žodžių ilgius tyrė ir William Palin Elderton (1877–1962), kurio tyrimo objektas buvo anglų kalbos žodžių ilgis (Elderton, 1949; Grzybek, 2007). O Shannon ir Weaver (1949) kalbos tyrimams pritaikė informacijos teoriją. Jų darbai buvo skaičiavimų, susijusių su įvairiomis kalbos savybėmis, tokiais kaip entropija ir kalbinis perteklius (angl.

redundancy), pradžia, nors ir neleido gauti gilesnių įžvalgų apie kalbą, kadangi šiuose skaičiavimuose neatsižvelgiama į lingvistinę semantiką. Vis dėlto, entropija ir kalbinis perteklius ir šiandien taikomi kaip tam tikros kalbos ar teksto charakteristikos.

Galiausiai Gustav Herdan (1897–1968) prie KL raidos prisidėjo įvairių sričių tyrimais. Jis tyrė anoniminės autorystės nustatymo problemas, atliko tyrimų stilometrijos, kalbos kitimo, kalbų maišymosi, ryšio tarp skirtingų žodžių ir visų žodžių tekste, žodžio ilgio ir žodžių pasikartojimų dažnumų pasiskirstymo, ryšio tarp teksto ilgio ir žodyno apimties bei teksto ilgio ir stiliaus, tyrė kalbų tipologiją (Best, 2007). Publikuotuose darbuose Herdan pristatė keletą kalbos dėsnų, tarp jų – Zipf'o¹⁹ ir Zipf'o-Mandelbrot'o skirstinio²⁰, Puasono skirstinio²¹ ir log-normaliojo skirstinio²². Tad Herdan laikomas vienu svarbiausių KL pradininkų (Best ir Altmann, 2018).

1.3.4 KL plėtros (nuo 1970 m.) ir Köhler sinergetinės lingvistikos (nuo 1993 m.) laikotarpiai

Gabriel Altmann (nuo 1970 m. iki 2020 m.) laikomas svarbiausiu tyrėju KL plėtros laikotarpiu nuo 1970 m. Iš tikrųjų, negalima pasakyti, kad šis plėtros laikotarpis yra pasibaigęs, KL toliau vystoma, taikomi metodai gana sparčiai kinta. Tad galima teigti, kad šis plėtros etapas tebesitęsia. Grįžtant prie Altmann, prie KL raidos jis prisidėjo hipotezių apie kalbos dėsnius kėlimu, matematiniu formulavimu ir empiriniu testavimu (Altmann, 1985; Altmann, 1993; Grzybek ir Köhler, 2011; Altmann ir Köhler, 2015). Daugelis kalbos dėsnų, tarp jų – Menzerath-Altman²³ (Altmann, 1980) ir Piotrowski'o²⁴

¹⁹ Apie Zipf'o dėsnį plačiau žr. 1.4.1 sk.

²⁰ Zipf'o-Mandelbrot'o skirstinio – apibendrintasis Zipf'o dėsnis (Silagadze, 1999). Pavadintas šį dėsnį apibendrinusio matematiko Benoit Mandelbrot vardu.

²¹ Tikimybių pasiskirstymo dėsnis, pavadintas jį pirmąsyk aprašiusio matematiko Siméon Denis Poisson vardu (Rothschild, 1986).

²² Log-normalusis skirstinys – tikimybių pasiskirstymo dėsnis. Galioja, kai kintamieji gali įgyti tik teigiamas reikšmes, o atlikus logaritminę transformaciją jo skirstinys tampa normalusis (Rothschild, 1986).

²³ Menzerath-Altman dėsnis nusako ryšį tarp lingvistinio konstrukto dydžio ir jį sudarančių dalių dydžių, pavyzdžiui, kuo ilgesnis žodis, tuo trumpesni jį sudarantys skiemenys arba kuo ilgesnis sakinytis, tuo trumpesnės jį sudarančios frazės ir panašiai (Mačutek ir Mikros, 2015).

²⁴ Piotrowski'o dėsnis nusako naujų kalbinių formų sklaidą kalbinėje bendruomenėje, t. y. tai yra kalbos kitimo dėsnis (Köhler ir Naumann, 2016).

(Altmann, 1983) dėsniai, jo dėka įgavo modernias formas, kitaip sakant – buvo matematiškai formalizuoti, jiems sukurtos matematinės formulės.

Altmann ir jo kolegas mokslininkus domino KL metodų taikymas atskirose kalbos tyrimų srityse (tokiose kaip fonemika (Bernhard ir Altmann, 2008; Nemcová ir Altmann, 2008; Altmann, 2016), semantika (Strauss ir Altmann, 2003; Fan ir Altmann, 2008; Kelih ir Altmann, 2015), teksto analizė (Altmann ir Köhler, 2015; Zörnig ir Altmann, 2016; Roelcke, Popescu ir Altmann, 2017), leksikologija (Best ir Altmann, 1986; Altmann, 2002a), ir kitose), bet ypač – kalbos dėsniai (Altmann, 1980; Altmann, 2002b; Wimmer ir Altmann, 2007), kalbos teorija (Wimmer ir Altmann, 1995; Popescu ir Altmann, 2014; Altmann, 2016) bei (įvairių kalbos vienetų) tikimybiniai pasiskirstymai (Popescu, Best, ir Altmann, 2014; Andreev, Popescu ir Altmann, 2017; Pelegrinova ir Altmann, 2017).

Kartu su Altmann paminėtinas ir Peter Grzybek (1957–2019), prie KL plėtros prisidėjęs žodžių ilgių (Fan, Grzybek, ir Altmann, 2010; Grzybek, 2016) bei grafemų ir fonemų pasikartojimų dažnumų slavų kalbose tyrimais (Grzybek, 2005; Grzybek ir Rusko, 2009). Su Altmann ir Grzybek KL interesais dalijosi ir Luděk Hřebíček (1934–2015). Jis domėjosi kalbų tipologija ir turkų kalbos sintakse. Remdamasis turkiškų tekstų pavyzdžiais jis parodė Menzerath'o dėsnio veikimą, t. y. kalbos lygmenų struktūravimą nuo smulkiausio iki stambiausio (Hřebíček, 1996). Hřebíček taip pat sukūrė koherentiškumo požymių tekste pasiskirstymo modelį ir jį patikrino su tektais turkų kalba (Hřebíček, 1986). Be to, anot šio tyrėjo, *hreb*'ai (agreguotos sekos, kurias sudaro pagrindinis žodis ir visi į jį nurodantys žodžiai) taip pat paklūsta pasiskirstymo kalbos dėsniams, tik kompleksiškesniu lygmeniu (Tatar, Lupea, ir Altmann, 2014).

Tad KL plėtros laikotarpis, kuris, galima sakyti, tebesitęsia ir šiandien, išsiskiria hipotezių apie įvairius kalbos dėsningumus kėlimu, bandymais šiuos dėsningumus užrašyti matematine kalba bei empiriniu hipotezių testavimu, siekiant ištirti kalbos dėsningumų universalumą. Po šio laikotarpio, ar, tiksliau, su šiuo laikotarpiu persidengia naujausias sinergetinės lingvistikos laikotarpis, kuriame siekiama paskirus kalbos dėsningumus sujungti į vieningą sistemą.

Reinhard Köhler reprezentuoja šį naujausią KL laikotarpį ir yra pagrindinis šios krypties tyrėjas. Köhler atlieka tyrimus kiekybinės ir sisteminės-teorinės lingvistikos srityse. Jis kiekybinę lingvistiką pastūmėjo į priekį vystydamas tarpusavyje susijusių kalbos dėsnų sistemą, kurios tikslas – suprasti, paaiškinti ir prognozuoti kalbos savybes, fenomenus ir procesus, t. y. tapdamas sinergetinės lingvistikos pradininku (Köhler, 2005b; Köhler, 2007). Jau minėtas Altmann, nors ir laikomas KL plėtros laikotarpio atstovu,

sėkmingai bendradarbiavo ir Köhler. Tad, kaip jau buvo minėta, KL istorijos etapų ribos yra apytikslios, negriežtos.

Sinergetinės lingvistikos tyrimai, atlikti Köhler, Altmann ir jų kolegų tyrėjų, apima, pavyzdžiui, kalbos fenomenų modeliavimą (Altmann ir Köhler, 1996; Tamaoka ir Altmann, 2005), ryšių tarp įvairių kalbos charakteristikų radimą (pavyzdžiui, tarp polisemijos ir žodžio ilgio (Miangah ir Altmann, (2016), žodžio ilgio ir žodžio dažnio (Strauss, Grzybek, Altmann, 2007), žodžių tvarkos sakinyje ir kalbos dalių sistemos ypatumų (Vulanović ir Köhler, 2009)), kalbos elementų sekų dėsningumų tyrimus (Köhler, 2015) (pavyzdžiui, kalbos dalių pasikartojimo (Beliankou ir Köhler, 2010; Köhler ir Naumann, 2016), pasikartojančių elementų (tokių, kuriems būdingas pastovus ilgis, didėjantis ilgis ar politekstualumas²⁵) sekų tekste tyrimus (Köhler ir Naumann, 2008), skirtingų kalbos lygmenų vienetų ir struktūrų (sistemų) tyrimus (pavyzdžiui, gramatikos (Köhler ir Altmann, 2000; Köhler, 2003; Vulcanovic ir Köhler, 2005), leksikos (Köhler, 2005c)) ir kitus. Tad, kaip galima matyti, sinergetinės lingvistikos siekis – prognozuoti ir paaiškinti kalbos savybes ir joje vykstančius procesus, o galutinis tikslas – kalbos dėsningumų sujungimas į vieningą sistemą.

KL istorijoje įvairių disciplinų tyrėjai, daugiausiai matematikai, kūrė ir taikė matematinius modelius bei vystė kitus kiekybinius metodus lingvistinių problemų sprendimui. Šiandien KL galima laikyti brandžia disciplina, kuriai būdingas platus tyrimų spektras. Populiariausi tie metodai ir tos teorijos, kurie remiasi ne tik matematinėmis, bet ir kalbos žiniomis (pavyzdžiui, Benoit Mandelbrot' o darbai (1959; 1961)). Kitaip tariant, plačiausiai taikomi tie metodai, tos kiekybinės kalbos charakteristikos, kurie turi ne tik matematinę išraišką, bet ir interpretaciją iš kalbos perspektyvos. Šioje disertacijoje pristatomas tyrimas paremtas KL metodais, o FS tyrimui naudoti kiekybiniai indikatoriai pasirinkti ne tik todėl, kad jų vertės ne tik galima suskaičiuoti, jie turi matematinės formules, bet ir todėl, gautas skaitines vertes galima paaiškinti, ką jos reiškia kalbos požiūriu, taip įgyjant naujų žinių apie kalbą, o tiksliau – apie lietuvių kalbos FS. Tyrime naudoti kiekybiniai indikatoriai, jų ypatybės pristatomi tolesniame poskyryje.

²⁵ Politekstualumas – tai intertekstinės sąsajos (Versekenaitė, 2008), t. y. tekstui įtaką turėjusių kūrinių, autorių, idėjų ir pan. tinklas.

1.4 Kiekybiniai indikatoriai FS kalbinės raiškos tyrime

Disertacijoje pristatomas tyrimas remiasi Miroslav Kubát (2016), Peter Zörnig ir kolegų (2016), Ioan-Iovitz Popescu ir kolegų (Popescu et al., 2010; Popescu, Čech ir Altmann, 2011a; Popescu, et al., 2017), Radek Čech (2011; 2016), Čech ir kolegų (Čech, Popescu ir Altmann, 2014; Čech, Garabík, ir Altmann, 2015), Kubát ir Jiří Milička (2013) darbais. Iš šių darbų lietuvių kalbos administracinio, mokslinio ir publicistinio FS kalbinės raiškos tyrimui pasirinkti 9 indikatoriai:

1. vidutinis žodžio / žodžio formos ilgis (VŽFI),
2. indikatorius a ,
3. žodžių ir žodžių formų įvairovės indikatorius R_1 ,
4. žodžių ir žodžių formų koncentruotumo indikatorius RR_{mc} ,
5. informacijos srauto įvertis MATTR (angl. *Moving Average Type Token Ratio*),
6. teminės koncentracijos įvertis TK,
7. teksto aktyvumo / dinamiškumo indikatorius A,
8. teksto deskriptyvumo / aprašomumo indikatorius D,
9. tarpveiksmožodinių atstumų indikatorius VA.

Be to, kad šie indikatoriai žymi skirtingus kalbinės raiškos aspektus (plačiau žr. 1.4.2 ir 1.4.3 sk.), dar viena jų pasirinkimo priežastis buvo tai, kad jie yra nepriklausomi arba beveik nepriklausomi nuo teksto ilgio (Kubát, 2015, 127; Kubát ir Čech, 2016b; Zörnig, Kelih ir Fuks, 2016). Kitaip sakant, šių indikatorių vertėms teksto ilgis įtakos neturi arba beveik neturi. Taip pat, remiantis kitų autorių tyrimais, kuriuose buvo naudojami kiekybiniai indikatoriai, buvo iškelta prielaida, kad pasirinktieji indikatoriai turėtų leisti skirti tiriamuosius FS. Indikatoriai a , R_1 , RR_{mc} , MATTR bei iš dalies TK yra paremti teksto dažnine struktūra²⁶, kurią apibūdina Zipf'o dėsnis, t. y. tekstų savybės, kurias žymi šie indikatoriai, yra matematiškai išvestos remiantis Zipf'o dėsniu. Indikatoriai VŽFI, A, D ir VA remiasi kitais statistiniais dėsningumais.

²⁶ Žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu ir Altmann, 2015).

1.4.1 Zipf'o dėsnis

Šiame skyriuje trumpai pristatomas Zipf'o dėsnis, kadangi 5 (a , R_1 , RR_{mc} , $MATTR$ ir iš dalies TK) disertacijoje pristatomame tyrime naudojamų indikatorių savybės yra išvestos (matematiškai) remiantis šiuo dėsniu. Tad šis dėsnis aptariamas arčiau minėtų indikatorių, o ne 1.3 sk. „Kiekybinė lingvistika: trumpa istorija“. Zipf'o dėsnį aprašė amerikiečių lingvistas Džordžas Kingslis Zipf'as (1935, 1949 m.), tyrinjęs statistines daugelio kalbų savybes (Powers, 1998). Zipf'o dėsnis teigia, kad esama matematinės priklausomybės tarp žodžio dažnio ir jo vietos dažniniame sąrašė, taip pat tarp žodžio dažnio ir žodžių, turinčių tą dažnį, skaičiaus (Altmann ir Gerlach, 2016). Tad Zipf'o dėsnis nusako teksto dažninę struktūrą (Čech, 2015; Popescu ir Altmann, 2015). Vėlesni tyrimai parodė, kad daugeliui kalbos fenomenų, pavyzdžiui, fonemų ir raidžių grandinėlių dažniams, gramatinėms konstrukcijoms, formulinėms frazėms (angl. *formulaic phrases*), etc., šis dėsnis taip pat galioja (Gries ir Ellis, 2015).

Zipf'o dėsnis paprastai aiškinamas „mažiausių pastangų principu“: žodžiai tampa ilgesni leidžiantis dažniniu sąrašu tolyn, o tai susiję su kalbos ekonomiškumu, nes jei dažnos žodžių formos būtų ilgos, kalba taptų neefektyvi ir komunikavimas reikalautų daug energijos (Powers, 1998; Utkā, 2004, 38—39). Taigi žodžių dažniai yra reikšmingi tiek žmonių komunikacijai, tiek kalbos technologijoms bei kalbos išmokimui (Piantadosi, 2014) ir juos, t. y. žodžių dažnius, galima charakterizuoti minėtu paprastu matematiniu dėsniu (oficialią matematinę formulę žr. 1.3.2 sk.): dažnis * rangas (t. y. pozicija dažnių lentelėje) = konstanta (Bružaitė ir Rekašius, 2016). Kitaip sakant, žodžio ar žodžio formos dažnį ir jo poziciją dažninėje lentelėje sieja pastovus skaitinis dydis. Tad Zipf'o dėsnis leidžia prognozuoti tam tikrų kalbos vienetų (žodžių formų, lemų, pastoviųjų žodžių junginių ir pan.), turinčių tam tikrą savybę (pavyzdžiui, ilgumą-trumpumą, t. y. ilgi / trumpi žodžiai), skaičių arba, kitaip, jų tikimybinį pasiskirstymą. Šios disertacijos kontekste Zipf'o dėsnis aktualus, nes tyrime naudoti indikatoriai a , R_1 , RR_{mc} , $MATTR$ ir iš dalies TK yra paremti teksto dažnine struktūra, kurią šis dėsnis apibūdina.

1.4.2 Teksto dažnine struktūra paremti kiekybiniai indikatoriai

Kaip jau minėta 1.4 sk. pradžioje, šioje disertacijoje aprašomame tyrime naudojami 5 teksto dažnine struktūra paremti arba iš dalies paremti

indikatoriai – a , R_1 , RR_{mc} , $MATTR$ ir TK . Jie bei jų savybės pristatomi šiame skyriuje, kartu pateikiant skaičiavimo formules ir verčių tiriamiesiems tekstams AST , MST ir PST pavyzdinius fragmentus.

1.4.2.1 Indikatorius a

Indikatorius a yra nuo teksto ilgio nepriklausomas h -taško variantas (Zörnig, Kelih ir Fuks, 2016). Indikatorius h -taškas žymi poziciją žodžių dažnių lentelėje, kur žodžio rangas (vieta žodžių dažnių lentelėje) yra lygus jo dažniui (Popescu, 2009, 17). Kalbos požiūriu, šis indikatorius tekstą sudarančių žodžių dažninėje lentelėje žymi ribą tarp labai dažnų žodžių ir žodžių formų (paprastai tai būna funkciniai žodžiai ir jų formos) ir retesnių žodžių ir jų formų (dažniausiai tai būna žodžiai, priklausantys pagrindinėms kalbos dalims). Funkciniais žodžiais laikomi prielinksniai, dalelytės, jungtukai (Utkā, 2004, 36) bei įvardžiai (Utkā, 2005, 51). Pagrindinėms kalbos dalimis laikomi daiktavardžiai, būdvardžiai, skaitvardžiai, veiksmažodžiai,rieveksmiai (Utkā, 2004, 36–37). Kitaip tariant indikatorius h -taškas žymi apytikslią ribą, kur tekstą sudarančių žodžių dažninėje lentelėje baigiasi labai dažni (paprastai – funkciniai) žodžiai ir prasideda retesni (priklausantys pagrindinėms kalbos dalims) žodžiai. Indikatoriaus vertės skaičiuojamos pagal šią matematinę formulę (Popescu, 2009, 17):

$$r = f(r) \quad (2)$$

r žymi žodžio rangą/poziciją dažninėje lentelėje, o $f(r)$ žymi žodžio, kuriam priklauso rangas/pozicija r dažninėje lentelėje, dažnumą.

Kaip jau minėta, h -taško skaitinė reikšmė yra priklausoma nuo teksto ilgio, o tai tyrėjui ne visada gali būti paranku. Tad Ioan-Iovitz Popescu (2009, 19) h -tašką normalizavo, sukurdamas indikatorių a , skaičiuojamą pagal šią formulę:

$$a = \frac{N}{h^2} \quad (3)$$

N žymi analizuojamo teksto ilgį žodžiais ir jų formomis, o h – h -tašką.

Indikatorius a (kitai – normalizuotas h -taškas) dalina tekstą į 2 dalis, kurios yra adaptuotos pagal teksto ilgį žodžiais (Zörnig, Kelih ir Fuks, 2016). Kitaip tariant, indikatorius a atvaizduoja dažninės lentelės sritį, esančią virš h -taško, t. y. tą dažninės lentelės dalį, kurią sudaro labai dažni (paprastai – funkciniai) žodžiai ir jų formos. Kuo didesnė a vertė, tuo tekstą sudarančių

žodžių dažninėje lentelėje labai dažnų žodžių sritis yra mažesnė, o kuo indikatorius vertė mažesnė – tuo ši sritis yra didesnė. Kitaip tariant, kuo didesnė indikatorius *a* vertė, tuo analizuojamą tekstą sudarančių žodžių dažninėje lentelėje labai dažni (paprastai funkciniai) žodžiai baigiasi arčiau dažninės lentelės viršaus, t. y. tuo analizuojamo teksto mažesnę dalį sudaro labai dažni žodžiai ir žodžių formos, o kuo šio indikatorius vertė mažesnė – tuo labai dažni žodžiai ir žodžių formos baigiasi toliau nuo dažninės lentelės viršaus, kitaip – tuo didesnę teksto dalį sudaro labai dažni žodžiai ir žodžių formos.

2 lentelė. Indikatorius *a* verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams.²⁷

	Tekstas	Tekstynas	Indikatorius <i>a</i>
1	adm113-28	AST	1,58
2	adm142-38	AST	2,43
3	adm261-75	AST	2,44
4	adm198-21	AST	2,55
5	adm83-1	AST	2,59
6	20150811_1936_delfi-veidai	PST	37,00
7	20160317_0831_delfi-veidai	PST	37,00
8	adm105-328	AST	39,50
9	20160926_1103_delfi-gyvenimas	PST	39,78
10	20150619_1150_delfi-gyvenimas	PST	43,22

2-oje lentelėje pateiktos 5 mažiausios ir 5 didžiausios indikatorius *a* vertės. Kaip galima matyti šioje lentelėje, mažiausios *a* vertės priklauso AST tekstams. Tai rodo, kad labai dažni žodžiai ir jų formos (paprastai – funkciniai, bet nebūtinai) yra tipiškai šiems tekstams. Kita vertus, 4 iš 5 tekstų, turinčių didžiausias *a* vertes (žr. 3 lentelės 5-10 eilutes), priklauso publicistiniam stiliui, Veidų ir Gyvenimo teminėms kategorijoms. Trečia pagal didumą indikatorius vertė (žr. 2 lentelės 8-ą eilutę) priklauso administracinio stiliaus tekstui. Šiems penkiems tekstams būdinga tai, kad jiems būdingi pasikartojimais retesni žodžiai ir jų formos.

²⁷ Pateiktos indikatorius *a* vertės yra tik pavyzdžiai, jos negali būti naudojamos FS pagrindimui.

1.4.2.2 Indikatorius R_1

Indikatorius R_1 paprastai naudojamas žodyno turtingumui matuoti (Kubát, Matlach ir Čech, 2014, 14). Dažniausiai žodyno turtingumas vertinamas skirtingų žodžių skaičiumi tekste (Hoover, 2003). Lietuvių kalbai būdinga kaitybinių formų gausa, todėl ši žodyno turtingumo traktuotė nėra tinkama, nebent būtų tiriamos ne žodžių formos, o lemos, t. y. žodžių antraštinės formos (Rimkutė, 2006). Kadangi šioje disertacijoje funkcinų stilių kalbinės raiškos požymiams analizuoti pasitelktos kaitybinės žodžių formos, o ne lemos, šiuo atveju laikoma, kad R_1 yra žodžių ir jų formų įvairovės indikatorius. R_1 skaičiuojamas pagal šią matematinę formulę (Kubát, Matlach ir Čech, 2014; Zörnig, Kelih ir Fuks, 2016):

$$R_1 = 1 - \left(F(h) - \frac{h^2}{2N} \right) \quad (4)$$

$F(h)$ žymi sukaupąjį santykinį dažnį iki h -taško (žr. 1.4.2.1 sk.), h žymi h -tašką, N žymi teksto dydį, t. y. jų sudarančių žodžių formų skaičių.

R_1 vertinama žodžių ir jų formų įvairovė – tai iš esmės retesnių žodžių (šieji paprastai priklauso pagrindinėms kalbos dalims) proporcijos analizuojamame tekste įvertinimas (Zörnig, Kelih ir Fuks, 2016; Xiao ir Sun, 2018). Kitaip sakant, kuo didesnė indikatoriaus vertė (galimos vertės pasiskirsto nuo 0 iki 1), tuo analizuojamam tekstui būdingesnė pasikartojimų atžvilgiu retesnių žodžių ir jų formų, gausa.

3 lentelė. Indikatoriaus R_1 verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams²⁸

	Tekstas	Tekstynas	Indikatorius R_1
1	adm35-27	AST	0,240
2	adm9-16	AST	0,321
3	adm46-59	AST	0,324
4	adm228-11	AST	0,342
5	adm35-28	AST	0,372
6	20150701_1045_delfi-auto	PST	0,968
7	20160719_1722_delfi-auto	PST	0,969

²⁸ Pateiktos indikatoriaus R_1 vertės yra tik pavyzdžiai, jos negali būti naudojamos FS pagrindimui.

8	20160502_1326_delfi-auto	PST	0,970
9	20140529_1308_delfi-auto	PST	0,973
10	20150322_0916_delfi-auto	PST	0,973

3-oje lentelėje pateikiamos 5 didžiausios ir 5 mažiausios indikatoriaus R_1 . Kaip galima matyti, mažiausios R_1 vertės administracinio stiliaus tekstams, o didžiausios – publicistinio. Tad pirmiesiems tekstams būdinga, lyginant su kitais 3-oje lentelėje pateiktais tekstais, maža žodžių ir jų formų įvairovė, t. y. nebūdinga retesnių, paprastai pagrindinėms kalbos dalims priklausančių, žodžių ir jų formų gausa. Antrajai tekstų grupei, priešingai, būdinga didelė žodžių ir jų formų įvairovė – didelė pasikartojimais retesnių žodžių ir jų formų gausa.

1.4.2.3 Indikatorius RR_{mc}

Indikatorius RR_{mc} yra normalizuotas indikatorius RR. Indikatorius RR (angl. *repeat rate*) paprastai naudojamas teksto žodyno koncentracijai įvertinti (Popescu, 2009, 166). Kitaip tariant, kuo tekste yra mažesnis labai dažnų žodžių ir žodžių formų skaičius su itin dideliais pasikartojimų dažniais, tuo laikoma, kad teksto žodynas yra koncentruotesnis. Disertacijoje atliekamo tyrimo atžvilgiu RR tikslingiau laikyti žodžių ir jų formų koncentracijos tekste įverčiu, nes indikatoriaus verčių skaičiavimui naudojamos žodžių formos, o ne lemos. RR skaičiuojamas pagal šią matematinę formulę (Popescu, Mačutek ir Altmann, 2009, 167):

$$RR = \sum_{r=1}^V p_r^2 \quad (5)$$

r žymi žodžio rangą/poziciją analizuojamą tekstą sudarančių žodžių ir jų formų dažnių lentelėje, V žymi analizuojamą tekstą sudarančių žodžių ir jų formų tipų skaičių (t. y. skirtingų žodžių ir jų formų skaičių), p_i žymi individualias žodžių formų analizuojamame tekste pasikartojimų tikimybes, apskaičiuotas pagal santykinius žodžių formų dažnius tame analizuojamame tekste.

RR_{mc} (McIntosh, 1967) yra santykinis RR, tinkamesnis nei originalusis indikatoriaus verčių lyginimui su kitų indikatorių vertėmis (Kubát, Matlach ir Čech, 2014, 19), todėl jis ir buvo pasirinktas šios disertacijos tyrimui kaip vienas iš FS kalbinę raišką žyminčių indikatorių. RR_{mc} kaip ir originalusis RR,

šios disertacijos tyrimo atveju žymi tekste vartojamų žodžių ir jų formų koncentruotumą (Kubát, Matlach ir Čech, 2014, 16; Popescu, et al., 2017), t. y. kuo aukštesnė šio indikatorius reikšmė, tuo tekstui būdingesnė didesnė žodžių ir jų formų koncentracija – analizuojamam tekstui būdingi ji sudarančių žodžių ir žodžių formų pasikartojimų gausa. RR_{mc} vertinamo tekste vartojamų žodžių koncentruotumo reikšmės pasiskirsto intervale nuo 0 iki 1, o tai patogu, kai norima šio indikatorius raišką palyginti su kitais indikatoriais. RR_{mc} yra normalizuotas RR , labiau tinka lyginimui su kitais indikatoriais. RR_{mc} skaičiuojamas pagal šią matematinę formulę (Popescu, Čech ir Altmann, 2011a):

$$RR_{mc} = \frac{1 - \sqrt{RR}}{1 - \frac{1}{\sqrt{V}}} \quad (6)$$

V žymi žodžių formų tipų skaičių, RR – originalųjį indikatorius RR .

4 lentelė. Indikatorius R_1 ir RR_{mc} verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams²⁹

	Tekstas	Tekstynas	Indikatorius RR_{mc}	Indikatorius R_1
1	adm213-177	AST	0,740	0,474
2	adm228-11	AST	0,771	0,342
3	adm46-59	AST	0,771	0,324
4	adm9-16	AST	0,782	0,321
5	adm266-151	AST	0,808	0,451
6	20150217_1437_delfi-mokslas	PST	0,991	0,960
7	20160919_1552_delfi-auto	PST	0,991	0,964
8	20160502_1326_delfi-auto	PST	0,991	0,970
9	20150322_0916_delfi-auto	PST	0,991	0,973
10	20140920_1423_delfi-auto	PST	0,992	0,975

²⁹ Pateiktos indikatorius R_1 ir RR_{mc} vertės yra tik pavyzdžiai, jos negali būti naudojamos FS pagrindimui.

4-ioje lentelėje pateiktos 5 didžiausios ir 5 mažiausios RR_{mc} reikšmės iš visų tiriamų tekstų. Kaip galima matyti, mažiausia žodžių ir jų formų koncentracija šioje lentelėje būdinga AST tekstams, o didžiausia – PST tekstams. Kitaip tariant, iš lentelėje pateiktų tekstų, administraciniams tekstams nėra būdingi žodžių ir jų formų pasikartojimai, o publicistinio stiliaus tekstams, priešingai, žodžių ir jų formų pasikartojimų gausa yra tipiška.

Indikatoriai R_1 ir RR_{mc} lingvistiniu požiūriu iš esmės yra vienas kitam priešingi, t. y. kuo didesnė žodžių ir jų formų įvairovė (didesnė R_1 vertė) būdinga tekstui, tuo šiam tekstui tipiškesnė mažesnė žodžių ir jų formų koncentracija (mažesnės RR_{mc} vertės). Kita vertus, šie indikatoriai skiriasi matematinio požiūriu: indikatorius R_1 vertina tik retesnių (t. y. paprastai priklausančių pagrindinėms kalbos dalims) žodžių ir jų formų įvairovę, tačiau neatsižvelgia į labai dažnus (paprastai – funkcinis) žodžius ir jų formas; RR_{mc} vertina visų žodžių ir jų formų koncentraciją analizuojamame tekste, t. y. tiek labai dažnų, tiek retesnių. Tad kaip galima matyti 4-ioje lentelėje, maža R_1 vertė nebūtinai reiškia, kad RR_{mc} vertė bus didelė. Administracinio stiliaus tekstų atžvilgiu tendencija, kai esant didelei RR_{mc} vertei, R_1 vertė yra maža, galioja, o publicistinio stiliaus atveju – ne (žr. 4 lentelę). Tad tyrimui pasirinkti abu indikatoriai.

1.4.2.4 Indikatorius MATTR

Indikatorius MATTR (angl. *Moving Average Type-Token Ratio*) yra modifikuotas iteracijos indeksas (angl. *Type-Token Ratio*; TTR), kurio vertės dėl modifikacijos nebėra, skirtingai nei TTR, priklausomos nuo teksto ilgio (Kubát ir Cech, 2016). Kitaip sakant, jei ilgesnio teksto TTR vertės yra didesnės, o trumpesnio – mažesnės, tai, naudojant MATTR, šios priklausomybės nebėra. MATTR atveju tekstas suskaidomas į pasirinkto ilgio atkarpas, kurių kiekvienai apskaičiuojamas TTR, o galutinis rezultatas, t. y. MATTR vertė, yra šių TTR vidurkis (Covington ir McFall, 2010). MATTR skaičiuojamas pagal šią matematinę formulę (Kubát ir Čech, 2016a):

$$\frac{\sum_{i=1}^{N-L} V_i}{N(N-L+1)} \quad (7)$$

N žymi teksto dydį – jį sudarančių žodžių ir jų formų skaičių, L – atkarpos, į kurias MATTR skaičiavimo metu suskaidomas tekstas, ilgis (žodžių ir jų formų skaičius), V_i – skirtingų žodžių ir jų formų skaičius atkarpoje L .

Anot I.-I. Popescu, iteracijos indeksas yra labiau informacijos srauto arba temos išdėstymo (angl. topic deployment) nei žodyno turtingumo/įvairovės matas, nors pastaroji interpretacija lingvistikos tyrimuose, ypač, kai tiriama anglų kalba, yra vyraujanti (Popescu, 2009, 233). Kaip teigia šis mokslininkas, informacijos srautas ilgame tekste yra didelis, o trumpame – mažas, t. y. tą pačią informaciją galima pateikti tiek 2000, tiek 20 žodžių. Sekant Popescu, šioje disertacijoje MATTR laikomas nuo teksto ilgio nepriklausomu informacijos srauto matu.

5 lentelė. Indikatorių MATTR ir R_1 verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams³⁰

	Tekstas	Tekstynas	Indikatorius MATTR	Indikatorius R_1
1	adm46-59	AST	14,759	0,324
2	adm228-11	AST	14,764	0,342
3	adm9-16	AST	14,946	0,321
4	adm35-167	AST	19,223	0,584
5	adm236-201	AST	20,858	0,414
6	20160629_1553_delfi-auto	PST	48,523	0,947
7	20150719_1236_delfi- mokslas	PST	48,526	0,960
8	20160908_2101_delfi- mokslas	PST	48,578	0,959
9	20160801_1721_delfi-auto	PST	48,594	0,952
10	20150903_1245_delfi-auto	PST	48,632	0,960

5-oje lentelėje pateikiamos 5 mažiausios ir 5 didžiausios MATTR vertės iš visų tiriamų tekstų. Lyginant šioje lentelėje pateiktų tekstų MATTR vertes galima matyti, jog AST tekstų jos yra mažiausios, o PST – didžiausios. Tad galima sakyti, jog publicistinio stiliaus tekstams būdingas didžiausias informacijos srautas (t. y. didžiausias kiekis skirtingų žodžių ir jų formų), o administracinio – mažiausias (t. y. mažiausias kiekis skirtingų žodžių ir jų formų).

Indikatoriai MATTR ir R_1 kalbos tyrimų kontekste turi panašumų, nes abu iš matematinės pusės yra susiję su teksto dažnine struktūra, kitaip tariant – analizuojamą tekstą sudarančių žodžių ir jų formų dažnių-rangų (pozicijų) pasiskirstymo dažninėje lentelėje ypatumais. R_1 , žymintis žodžių ir jų formų

³⁰ Pateiktos indikatorių MATTR ir R_1 vertės yra tik pavyzdžiai, jos negali būti naudojamos FS pagrindimui.

įvairovę, yra paremtas h-tašku, žyminčiu ribą tarp labai dažnų (paprastai – funkcinių, nors nebūtinai) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir jų formų arba tašką, kuriame žodžio formos dažnis yra lygus jo rangui (pozicijai) tekstą sudarančių žodžių ir žodžių formų dažniniame sąrašė. Nors abu indikatoriai vertina skirtingų žodžių ir jų formų kiekį tekste, tačiau rezultatai pasiekti naudojami skirtingi metodai. MATTR yra skirtingų žodžių ir visų žodžių santykio nustatyto ilgio teksto atkarpoms vidurkis, o R_1 iš esmės vertina žemiau h-taško esančių, t. y. retesnių pasikartojimais, žodžių ir jų formų įvairovę tekste. Didėjant MATTR vertėms, R_1 vertėms nebūtinai lygiai taip didėja (žr. 5 lentelės 4-5 ir 7-9 eilutes). Kadangi šie indikatoriai matematiškai skaičiuojami skirtingai, jie abu naudojami FS kalbinės raiškos tyrime.

1.4.2.5 Teminės koncentracijos indikatorius

Teminės koncentracijos (TK) indikatorius vertina, kiek teksto raiška sukoncentruota į gvildenamą temą (Kubát, Matlach ir Čech, 2014, 48; Čech, 2016; Kubát ir Čech, 2016b). Kiekviename tekste tema atsiskleidžia arba yra plėtojama turinio / teminių žodžių formomis, t. y. žodžių formomis, priklausančiomis pagrindinėms kalbos dalims – daiktavardžiams, būdvardžiams, skaitvardžiams, veiksmazodžiams, prievėksmiais (Utká, 2004, 36–37). Tad kuo tekstas temos atžvilgiu koncentruotesnis, tuo TK indikatoriaus vertė yra didesnė. Kalbant detaliau, teminė koncentracija reiškia, kad virš h-taško (nurodo ribą tarp labai dažnų ir retesnių žodžių ir jų formų) tarp labai dažnų (paprastai – funkcinių) žodžių formų yra įsiterpę ir turinio / teminių žodžių formų, t. y. pastarosios savo dažnumu nedaug nusileidžia funkciniams žodžiams.

Šie turinio žodžiai, būdami savotiška anomalija, laikomi teksto teminiais žodžiais, nusakančiais teksto temą (Kubát, Matlach ir Čech, 2014, 48–49). Tad tekstams, kurių TK vertės yra didelės, būdingos labai dažnos (t. y. dažnumu nedaug nusileidžiančios funkciniams žodžiams ir jų formoms, paprastai esančioms dažninės lentelės viršuje) turinio/teminių žodžių formos. Pavyzdžiui, moksliniams tekstams paprastai būdingesnė didesnė teminė koncentracija nei grožinės literatūros tekstams (Popescu, 2009, 100). Kitaip sakant, moksliniams tekstams, lyginant su grožinės literatūros tekstais, būdingesni turinio / teminiai žodžiai ir žodžių formos, savo dažnumu bemaž prilygstantys funkciniams žodžiams.

Paminėtina, kad teminės koncentracijos indikatorius yra itin jautrus trumpiems tekstams (Kubát ir Čech, 2016b), t. y. labai trumpų tekstų teminės

koncentracijos rodiklis yra itin aukštas. Tad į FS tyrimui sudarytus tikslinius tekstynus įtraukti ne trumpesni nei 200 žodžių ilgio tekstai. Pereinant prie skaičiavimo, TK yra teksto atskirų turinio žodžių, dažninėje lentelėje esančių virš h-taško, teminių svorių suma (Chen ir Liu, 2018). Turinio / teminio žodžio ar jo formos, esančios virš h-taško, teminis svoris apskaičiuojamas pagal šią formulę:

$$2 \frac{(h - r')f(r')}{h(h - 1)f(1)} \quad (8)$$

h žymi h-tašką, r' – teminio / turinio žodžio, esančio virš h-taško, pozicija (rangas) tekstą sudarančių žodžių ir jų formų dažninėje lentelėje, $f(r')$ – teminio / turinio žodžio, esančio virš h-taško, dažnis, $f(1)$ – dažniausio teksto žodžio ar žodžio formos dažnis.

Tada TK apskaičiuojama taip:

$$\sum_{r'=1}^T 2 \frac{(h - r')f(r')}{h(h - 1)f(1)} \quad (9)$$

6 lentelė. Teksto dažninės lentelės fragmentas

Pozicija pagal dažnumą	Žodis	Dažnumas
1	ir	30
2	streso	30
3	tai	18
4	kad	15
5	į	14
6	stresą	13
7	yra	13
8	su	12
9	sako	10
<u>10</u>	<u>iš</u>	<u>10</u>
11	laurinaitis	10
12	labai	7
13	budrys	7
14	stresu	7
15	mes	6

Toliau pateikiamas praktinis TK skaičiavimo pavyzdys. 6-oje lentelėje matomas tekstą sudarančių žodžių ir jų formų dažninės lentelės fragmentas. Joje h-taškas – ties 10-oje pozicijoje esančiu žodžiu iš (šio žodžio dažnis yra lygus jo rangui / pozicijai teksto dažninėje lentelėje). Virš šio žodžio randame 3 teminius / turinio žodžius: streso, stresą, sako. Kadangi žodžio būti formos dažnai yra sudėtinio tarinio dalis, remiantis „Dažniniu rašytinės lietuvių kalbos žodynu“ (Utkla, 2009, 1) 5 dažniausios formos (yra, buvo, būti, būtų, bus) nelaikomos teminiais žodžiais. Tad 6-oje lentelės pozicijoje esantis žodis yra tarp teminių žodžių nepatenka. Skaičiuojame teksto fragmento TK (ŽTS – žodžio teminis svoris):

$$\begin{aligned} & \check{Z}TS(streso) + \check{Z}TS(stresą) + \check{Z}TS(sako) \\ &= \left(\frac{2(10 - 2)30}{10(10 - 1)30} \right) + \left(\frac{2(10 - 6)13}{10(10 - 1)30} \right) \\ &+ \left(\frac{2(10 - 9)10}{10(10 - 1)30} \right) = 0,22 \end{aligned}$$

Sudėję tarp labai dažnų žodžių įsiterpusių 3 retesnių (turinio / teminių, priklausančių pagrindinėms kalbos dalims) teminius gauname, kad teksto fragmento TK lygus 0,22.

7 lentelė. Indikatoriaus TK verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams.

	Tekstas	Tekstynas	Indikatorius TK
1	20140505_1139_delfi-auto	PST	0,000
2	20140510_1054_delfi-auto	PST	0,000
3	20140513_1629_delfi-auto	PST	0,000
4	20140621_0837_delfi-auto	PST	0,000
5	20140703_1813_delfi-auto	PST	0,000
6	adm55-16	AST	1,850
7	adm174-189	AST	1,857
8	adm263-83	AST	1,875
9	adm6-91	AST	1,875
10	adm81-121	AST	1,878

7-oje lentelėje pateikiamos 5 didžiausios ir 5 mažiausios TK vertės iš visų disertacijoje tiriamų tekstų. Mažiausios vertės lentelėje priklauso PST tekstams, o didžiausios – AST. Kai TK vertė lygi 0, laikoma, kad tekstas

teminės koncentracijos atžvilgiu yra neutralus (Čech, Garabík ir Altmann, 2015). Kitaip sakant, tai reiškia, kad jam nebūdingi turinio / teminiai žodžiai ir jų formos, savo dažnumu nedaug nusileidžiantys funkciniam žodžiams, kurie paprastai tekste yra dažniausi. Tad lentelėje galima matyti, kad PST tekstai yra teminės koncentracijos atžvilgiu yra neutralūs, o administraciniai – labai koncentruoti, t. y. šiems tekstams būdingi turinio / teminiai žodžiai ir jų formos, savo dažnumu bemaž prilygstantys funkciniam žodžiams. Tai gali būti susiję su administracinio stiliaus standartiskumu ir šabloniškumu (Žilinskienė, 2002b; Vladarskienė, 2007; Župerka, 2012, 85; Smetonienė, 2013).

1.4.3 Kiti indikatoriai

Be teksto dažnine struktūra paremtų indikatorių, FS kalbinės raiškos tyrime naudojami 4 kiti statistiniai dėsniniai apibūdinimai nei apibūdina Zipf'o dėsnis, paremti indikatoriai: VŽFI (teksto sudėtingumo suprantamumo prasme įvertis (angl. *readability*)), A (teksto aktyvumo / dinamiškumo įvertis), D (teksto deskriptyvumo / aprašomumo įvertis) ir tarpveiksmažodinių atstumų įvertis VA. Šie indikatoriai apibūdina kitokias tekstų savybes nei teksto dažnine struktūra paremtieji, tad jie taip pat įtraukti į tyrimo naudojamų kiekybinių indikatorių rinkinį, o jų savybės, skaičiavimo formulės bei verčių tiriamiesiems tekstams AST, MST ir PST fragmentai pateikiami šiame skyriuje.

1.4.3.1 Vidutinis žodžių ir žodžių formų ilgis (VŽFI)

VŽFI indikatorius yra paprasta charakteristika, dažnai naudojama tekstų sudėtingumui suprantamumo prasme (angl. *readability*) įvertinti (van Oosten, Tanghe ir Hoste, 2010). Matematiškai indikatorius šioje disertacijoje atliekamo tyrimo kontekste vertina teksto žodžių formų ilgių aritmetinį vidurkį, kuris skaičiuojamas pagal šią formulę (Kubát, Matlach ir Čech, 2014, 47):

$$\frac{1}{N} \sum_{i=1}^N x_i \quad (10)$$

N žymi teksto ilgį, t. y. jį sudarančių žodžių ir jų formų skaičių, x_i žymi atskiros žodžio formos ilgį.

8 lentelė. Indikatoriaus VŽFI verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams

	Tekstas	Tekstynas	Indikatorius VŽFI
1	20160606_1155_delfi-gyvenimas	PST	5,013
2	20150430_0000_delfi-gyvenimas	PST	5,067
3	20151107_1753_delfi-gyvenimas	PST	5,146
4	20151209_2211_delfi-gyvenimas	PST	5,177
5	20160418_1746_delfi-gyvenimas	PST	5,180
6	adm57-85	AST	9,064
7	adm124-55	AST	9,069
8	adm250-18	AST	9,070
9	adm131-45	AST	9,196
10	adm131-25	AST	9,209

8-oje lentelėje pateikiamos 5 didžiausios ir 5 mažiausios indikatoriaus VŽFI vertės iš visų tiriamųjų tekstų. Mažiausios vertės lentelėje priklauso publicistinio FS tekstams, o didžiausios – administracinio stiliaus tekstams. Tai rodo, kad publicistiniams tekstams būdingi vidutiniškai trumpi žodžiai ir žodžių formos, o administraciniams tekstams – vidutiniškai ilgi. Kitaip tariant, publicistiniai tekstai, lyginant su administraciniais, suprantamumo atžvilgiu yra paprastesni. Administraciniai tekstai, lyginant su publicistiniais, priešingai, suprantamumo požiūriu yra sudėtingesni.

1.4.3.2 Aktyvumo (A) ir deskriptyvumo (D) indikatoriai

Aktyvumo (A) ir deskriptyvumo / aprašomumo (D) indikatoriai yra paremti idėja, kad teksto aktyvumą paprastai nusako veiksmažodžiai, o jam priešingą deskriptyvumą / aprašomumą – būdvardžiai (Kubát, Matlach ir Čech, 2014, 43; Kubát, 2016, 77–78). Kaip rodo M. Kubát čekų kalbos tekstų tyrimai, romano, pasakos ir panašių tekstų A įvertis bus didesnis (daugiau veiksmažodžių lyginant su būdvardžiais) nei mokslinio tyrimo ar kelionių aprašymo (šiuose ir panašiuose tekstuose sutinkama daugiau būdvardžių nei

veiksmazodžių) (Kubát, 2016, 80). Tad šie indikatoriai gali būti laikomi vienu iš būdų išmatuoti teksto „toną“, kitaip tariant – įvertinti tekstą per aktyvumo ir deskriptyvumo dimensijas (arba veiksmazodžių ir būdvardžių kiekį) (Zörnig ir Altmann, 2016). Indikatorius A skaičiuojamas pagal šią formulę:

$$\frac{\text{veiksm.}}{\text{veiksm.} + \text{būd.}} \quad (11)$$

Veiksm. žymi veiksmazodžių formų tekste skaičių, *būd.* – būdvardžių formų tekste skaičių. Indikatorius D yra priešingas indikatoriui A ir apskaičiuojamas pagal formulę:

$$1 - A \quad (12)$$

9 lentelė. Indikatorių A ir D verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams

	Tekstas	Tekstynas	Indikatorius A	Indikatorius D
1	adm213-212	AST	0,034	0,966
2	adm242-108	AST	0,047	0,953
3	adm58-1	AST	0,059	0,941
4	adm208-23	AST	0,067	0,933
5	adm64-94	AST	0,087	0,913
6	adm77-5	AST	1,000	0,000
7	adm95-32	AST	1,000	0,000
8	adm95-34	AST	1,000	0,000
9	adm95-58	AST	1,000	0,000
10	adm95-6	AST	1,000	0,000

9-oje lentelėje pateikiamos 5 didžiausios ir 5 mažiausios indikatorių A ir D vertės iš visų tiriamųjų tekstų. Tiek didžiausios, tiek mažiausios abiejų indikatorių vertės priklauso AST tekstams. Kadangi indikatoriai vienas kitam priešingi, kai vieno jų vertės didėja, kito – mažėja ir atvirkščiai. Didžiausios A (tuo pačių metu – mažiausios D) vertės priklauso Lietuvos Respublikos Prezidento dekretams dėl Lietuvos Respublikos pilietybės atsisakymo arba suteikimo. Šiuos tekstus iš esmės sudaro ilgas vardų, pavardžių, gimimo ir gyvenamosios vietovių sąrašas, kuriame yra vos keli veiksmazodžiai (pavyzdžiui, *suteikiu, įsigalioja*), bet nėra nei vieno būdvardžio.

Mažiausios A (ir kartu didžiausios D vertės) priklauso dokumentams, kurių turinys – taip pat sąrašas, tik kitoks. Teksto „adm213-212“, turinčio mažiausią iš lentelėje pateiktų A verčių, turinys – „Valstybinių studijų ir mokymo įstaigų bei valstybės pripažintų nevalstybinių švietimo įstaigų sąrašas“. Tad, nors šiame tekste, kaip ir Prezidento dekretuose dėl Lietuvos Respublikos pilietybės atsisakymo arba suteikimo, yra vos keli veiksmažodžiai (iš viso 4) jam būdingi būdvardžiai (iš viso 113), įeinantys į pavadinimus, pavyzdžiui, *Telšių aukštesnioji taikomosios dailės mokykla, Vilniaus lengvosios pramonės mokykla, Kauno taikomosios dailės mokykla* ir panašiai. Tad šiame kontekste didžiausios ir mažiausios A ir D vertės, priklausančios AST tekstams, galėtų būti labiau susijusios su šio stiliaus tikslingumu (A) (Kniūkšta, 2005; Pečkuvienė, 2013) ir tikslumu (D) (Vladarskienė, 2005; Pečkuvienė, 2013) negu su aktyvumu ir deskriptyvumu / aprašomumu.

1.4.2.3 Tarpveiksmažodiniai atstumai (VA)

VA indikatorius tekste vertina vidutinį atstumą (žodžių formomis) tarp dviejų iš eilės einančių veiksmažodžių (Kubát, Matlach ir Čech, 2014, 42). Galima sakyti, kad šis indikatorius apibūdina teksto sintaksinės struktūros kompleksiskumą (Kubát, 2016, 70–71), nors, žinoma toks vertinimas yra labai supaprastintas. Tarkime, turime itin paprastą tekstą, sudarytą iš trumpų, paprastų sakinių (tik veiksnio ir tarinio). Tokio teksto vidutinis atstumas tarp dviejų veiksmažodžių bus nedidelis. Kita vertus, sudėtingesniame tekste, sudarytame iš ilgesnių ir sudėtingesnių sakinių (pavyzdžiui, aukštųjų mokyklų mokomųjų tekstų (Bitinienė, 2007a)), vidutinis atstumas tarp dviejų iš eilės einančių veiksmažodžių būtų didesnis. VA skaičiuojamas tekstui, t. y. nekreipiama dėmesio į sakinius ir ribas tarp jų. Kuo teksto sintaksė sudėtingesnė, tuo VA vertė yra didesnė.

10 lentelė. Indikatoriaus VA verčių fragmentai tikslinių tekstynų AST, MST ir PST tekstams

	Tekstas	Tekstynas	Indikatorius VA³¹
1	20140416_1028_delfi-pilietis	PST	2,652
2	20140426_0900_delfi-pilietis	PST	2,658
3	20150222_1217_delfi-pilietis	PST	2,736
4	20141121_1539_delfi-gyvenimas	PST	2,771
5	20150915_2249_delfi-gyvenimas	PST	2,818
6	adm40-10	AST	398,333
7	adm69-17	AST	437,333
8	adm234-8	AST	440,600
9	adm198-10	AST	450,667
10	adm40-54	AST	463,000

10-oje lentelėje pateiktos 5 didžiausios ir 5 mažiausios VA vertės iš visų disertacijoje analizuojamų tekstų. Mažiausios vertės priklauso PST tekstams, o didžiausios – AST tekstams. Tai rodo, kad publicistiniams tekstams būdingesnė paprastesnė sintaksinė struktūra, o administraciniams – sudėtingesnė. Mažiausias VA vertes įgijo publicistiniai tekstai, priklausantys naujienų portalo Delfi *Piliečio* ir *Gyvenimo* teminėms kategorijoms. Pirmosios turinys – piliečių nuomonės, pasisakymai ir istorijos. Gyvenimo teminėje kategorijoje publikuojami su gyvenimo būdu susiję straipsniai. Abi teminės kategorijos susijusios su kasdienybe, paprastais, kasdieniais rūpesčiais ir problemomis, todėl galima daryti prielaidą, kad šie tekstai buvo rašomi gana paprasta, kalba su nesudėtinga sintakse. Mažiausias VA vertes įgijo Lietuvos Respublikos Prezidento dekretai dėl Lietuvos Respublikos pilietybės suteikimo arba atsisakymo. Šių tekstų ypatybė – ilgi vardų, pavardžių ir vietovių sąrašai, kuriuose randamos 3 veiksmažodžių formos: 1 – dokumento pradžioje (*suteikiu* arba *patenkinu*) ir 2 dokumento pabaigoje (*įsigalioja*, *prisiekus* arba *nurodyti*, *netenka*). Disertacijoje atliekamo tyrimo kontekste VA galima sieti, priklausomai nuo verčių didumo, su glaustumo, išsamumo, konkretumo raiška tekste.

³¹ Pateiktos vertės nurodo tekste vidutinį atstumą (žodžių formomis) tarp dviejų iš eilės einančių veiksmažodžių.

1.5 Tyrimų, kuriuose taikyti kiekybiniai indikatoriai tekstu savybėms tirti, apžvalga

Šioje disertacijoje lietuvių kalbos administracinis, mokslinis ir publicistinis FS tiriami pasitelkus įvairias kalbinės raiškos savybes žyminčius kiekybinius indikatorius. Tad šiame skyriuje apžvelgiami darbai, kuriuose taip pat taikyti skirtingi kiekybiniai indikatoriai įvairių su kalbos ar teksto analize susijusių uždavinių sprendimui. Paminėtina, kad, jei tai nėra tam tikro indikatoriaus savybių tyrimui skirtas kurio nors tyrėjo darbas, įvairioms kalbos savybėms tirti neretai pasitelkiamas ne vienas indikatorius, o jų rinkinys. Indikatorių pasirinkimui įtakos turi tirama problema, duomenų ypatumai³², indikatorių interpretacija kalbos atžvilgiu³³, kiti aspektai.

Kadangi kiekybinėje lingvistikoje taikomi metodai santykiškai greitai kinta, publikuojama nemažai metodologinio pobūdžio straipsnių. Indikatorių kontekste šiuose straipsniuose tiriamos jų, t. y. indikatorių, savybės, aptariamos interpretacijos kalbos požiūriu ir taikymo galimybės. Kitaip sakant, tirama, ką tiriamos su kalba ar tekstu susijusios problemos atžvilgiu rodo tam tikras indikatorius ar indikatorių rinkinys. Kiti tyrimai apima konkrečius indikatorių taikymo atvejus, t. y. jų naudojimą kalbos ir teksto analizei. Šiame poskyryje pristatomos abi tyrimų grupės. Paminėtina, kad stengtasi aptarti tik aktualius šioje disertacijoje kiekybinius indikatorius. Disertacijoje pristatomame tyrime nenaudoti indikatoriai trumpai paminimi, jei jie buvo įtraukti į apžvelgiamuose tyrimuose naudotų indikatorių rinkinį.

1.5.1 Metodologinio pobūdžio publikacijų apžvalga

Šiame poskyryje trumpai pristatomi aktualūs disertacijos kontekste metodologinio pobūdžio tyrimai, kuriuose naudojami kiekybiniai indikatoriai.

³² Pavyzdžiui, kadangi dalis indikatorių yra priklausomi nuo teksto ilgio, tiriamų tekstų heterogeniškumas jų ilgio / apimties atžvilgiu turi įtakos tyrėjo pasirinkimui.

³³ Indikatoriai skaičiuojami pagal aiškias formules, tačiau jų interpretacija, priklausomai nuo tiriamos problemos, gali būti nevienareikšmiška. Pavyzdžiui, indikatorius h-taškas matematiškai žymi tašką dažninėje lentelėje, kur žodžio dažnumas yra lygus jo rangui, t. y. pozicijai šioje lentelėje. Tiriant kalbų tipologijos problematiką, h-taško vertės taikomos kalbos analitiškumui-sintetiškumui įvertinti (Popescu, 2009, 23), o individualaus stiliaus ar žanrų tyrime gali būti naudojamas žodyno turtingumui (Zörnig, Kelih ir Fuks, 2016) ar tiesiog labai dažnų ir retesnių žodžių porcijų dydžiams tekste matuoti.

Šiuose tyrimuose analizuojamos indikatorių savybės. Nors mokslinėje literatūroje gausu kiekybinių indikatorių, tai, ką jie parodo kalbos atžvilgiu, bei kitos jų savybės nevisada būna pakankamai iširtos. Indikatorių savybių bei jų interpretavimo iširtumas turėjo įtakos pasirenkant indikatorius lietuvių kalbos FS kalbinės raiškos požymių kiekybiniam tyrimui.

Iš disertacijos tyrimui pasirinktų indikatorių bene daugiausiai savybių kontekste tiriamas teminės koncentracijos³⁴ indikatorius. Pavyzdžiui, buvo analizuojama šio indikatoriaus priklausomybė nuo teksto ilgio (Čech, Garabík ir Altmann, 2015). 20-ies tekstų slovakų kalba analizė parodė, kad trumpiems tekstams būdinga teminės koncentracijos verčių svyravimai, tad, autorių teigimu, šis indikatorius iš dalies priklausomas nuo teksto ilgio. Kita vertus, kitame teminės koncentracijos indikatoriaus taikymo anglų kalbai aspektų tyrime teigiama, kad teminė koncentracija ir antrinė teminė koncentracija³⁵ nepriklauso nuo tekstų ilgio, kai pastarųjų apimtis nuo 200 iki 6500 (Kubát ir Čech, 2016b). Dar viename teminės koncentracijos tyrime teigiama, kad teminės koncentracijos vertės priklauso nuo skirtingų žodžių skaičiaus tekste. Kadangi, kaip galima matyti, mokslininkai nesutaria dėl teminės koncentracijos priklausomumo nuo teksto ilgio, šioje disertacijoje remiamasi Čech, Garabík ir Altmann (2015) tyrimu ir laikomasi nuostatos, kad teminės koncentracijos indikatorius yra priklausomas nuo teksto ilgio tik esant labai trumpiems tekstams. Tam sudarant tikslinius lietuvių kalbos FS tekstynus į juos įtraukti tik ilgesni nei 200 žodžių tekstai. Kubát ir Čech (2016b) nustatyta viršutinė riba, disertacijos autorės manymu, nėra pakankamai pagrįsta, nes jų tiriamąją medžiagą kartu sudarė tekstų dalys (romanų skyriai), o ne tik ištisi tekstai.

Indikatorių savybių tyrimų kontekste paminėtinas indikatoriaus h-taško³⁶ tyrimas, kai tirtas indikatoriaus verčių pasiskirstymas lemuotiems ir nelemuotiems tekstams rusų, slovėnų ir ukrainiečių kalbomis analizė (Keliš, Rovenchak ir Buk, 2014). Tyrimas atskleidė, kad lemuotiems tekstams rusų ir

³⁴ Žymi, kaip tekstas yra sukonzentruotas į temą (Čech, Garabík ir Altmann, 2015; Kubát, 2015, 57–58). Plačiau žr. 1.4.2.5 sk.

³⁵ Teminės koncentracijos variantas (plačiau žr. 1.4.2.5 sk.). Neretai pasitaiko, kad tarp labai dažnų žodžių (paprastai – funkcinų) nebūna įsiterpę pagrindinėms kalbos dalims priklausančių žodžių, tam teminės koncentracijos vertė būna lygi 0 (Kubát, Matlach ir Čech, 2014, 52). Antrinė teminė koncentracija leidžia išspręsti šią problemą.

³⁶ Žymi tašką dažninėje lentelėje, kur žodžio ar žodžio formos dažnis yra lygus jo pozicijai toje lentelėje, o kalbos ir teksto tyrimų kontekste h-taškas žymi apytikslių ribų tarp labai dažnų (paprastai – funkcinų) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų (plačiau žr. 1.4.2.1 sk.).

ukrainiečių kalboms h-taško vertė yra mažesnė nei nelemuotiems, t. y. lemuotų tekstų riba tarp labai dažnų ir retesnių žodžių yra arčiau dažnės lentelės viršaus nei nelemuotų. Tai vaizduoja kalbų su turtinga morfologija ypatumus: lemuotuose tekstuose skirtingų žodžių skaičius yra mažesnis nei nelemuotuose, tačiau daugiau pirmuosiuose taip pat daugiau žodžių, turinčių didesnę dažnį. Dar vienoje metodologinio pobūdžio publikacijoje indikatoriai h-taškas ir žodžių dažnių-rangų kreivės ilgis³⁷ taikyti kalbos lygmenų (kategorijų) diversifikacijos tyrime (Popescu ir Altmann, 2008). Tyrime naudoti tekstai 20-ia kalbų. Tyrimo rezultatai parodė, kad fonologinių fenomenų diversifikacija yra mažiausia, semantinių – didžiausia, o morfologinių – viduryje tarp fonologinių ir semantinių (Popescu ir Altmann, 2008). Šios h-taško indikatorius savybės įdomios kalbos savybių kontekste, tačiau nors antrasis aptartas tyrimas ir neturėjo įtakos disertacijos tyrimo eigos pasirinkimams, pirmuoju pasiremta disertacijos tyrimui pasirenkant žodžių formas, o ne lemas.

Galusiai TTR gali būti skaičiuojamas ne tik žodžių lygmenyje bet ir, pavyzdžiui, įvairiems teksto segmentams: L-segmentams (tiriamųjų kalbos vienetų (morfemų, žodžių, sakinių ir pan.), kuriems būdingas pastovus arba didėjantis ilgis, sekoms), F-segmentams (tiriamųjų kalbos vienetų (morfemų, žodžių, sakinių ir pan.), kuriems būdingas pastovus arba didėjantis dažnis, sekoms) ir T-segmentams (tiriamųjų kalbos vienetų (morfemų, žodžių, sintaksinių konstrukcijų tipų ir pan.), kuriems būdingas pastovus arba didėjantis daugiareikšmiškumas, sekoms) (Köhler ir Naumann, 2008). Pasak straipsnio autorių, šie segmentai žodžių lygmenyje galėtų tikti tekstų tipologiniams tyrimams. Šis tyrimas disertacijos kontekste aktualus ateities FS kalbinės raiškos požymių kiekybinės analizės kontekste.

Šiame poskyryje trumpai pristatyti aktualūs disertacijos kontekste metodologinio pobūdžio tyrimai, kuriuose tiriamos indikatorių savybės. Indikatorių savybių ištirtumas yra svarbus pasirenkant indikatorius tyrimui. Jei indikatorius savybės nėra pakankamai ištirtos, gali kilti neaiškumų, kaip kalbos požiūriu paaiškinti gautus rezultatus. Apžvelgtais teminės koncentracijos ir h-taško savybių tyrimais buvo remtasi įtraukiant šiuos indikatorius į lietuvių kalbos FS kalbinės raiškos kiekybinę požymių analizę, o TTR taikymas įvairiems teksto segmentams įdomus ateities tyrimams.

³⁷ Žodyno turtingumas įvertis, paremtas dažnių-rangų (pozicijų) dažnėje lentelėje pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, 35–36; Juola, Mikros ir Vinsick, 2018).

1.5.2 Trumpa taikomojo pobūdžio literatūros apžvalga

Šiame poskyryje trumpai pristatomi aktualūs disertacijos kontekste taikomieji tyrimai, kuriuose naudojami kiekybiniai indikatoriai. Kadangi indikatoriai žymi skirtingas kalbos/teksto savybes, tyrimuose neretai naudojami indikatorių rinkiniai, todėl, jei įeina į tyrimo naudojamų indikatorių grupę, apžvalgoje paminimi ir disertacijos tyrime nenaudoti indikatoriai. Taikomuosius tyrimus, kuriuose naudoti kiekybiniai indikatoriai, apytiksliai galima suskirstyti į kelias grupes: retorikos tyrimus, tekstų grupavimo, remiantis įvairia logika (pavyzdžiui, pagal žanrus, tipus ir pan.), tyrimus, autorystės tyrimus, įvairių tipų tekstų analizės ir tipologinius tyrimus. Toliau šiame poskyryje trumpai pristatomi kiekybiniai tyrimai, pradedant retorikos tyrimais ir baigiant tipologiniais.

Čekijos prezidentų retorikos tyrimas, kuriame naudoti indikatorius λ ³⁸ (teksto dažninės struktūros³⁹ indikatorius) ir R_1 (žodyno turtingumo matas; plačiau žr. 1.4.2.2 sk.) atskleidė, kad pagal šiuos indikatorius, kitaip – žodyno turtingumą ir teksto dažninę struktūrą, prezidentų retorika yra panaši (Čech, 2011). Tai reiškia, kad šie du indikatoriai leido nustatyti, kad su pareigomis susijęs kalbėjimo stilius šiuo atveju nustelbė individualųjį stilių. Panašus tyrimas atliktas analizuojant JAV prezidentų inauguracines kalbas (Kubát ir Čech, 2016a). Prezidentų retorikos analizei taikyti indikatoriai MATTR, antrinė teminė koncentracija⁴⁰ ir aktyvumas⁴¹. Priešingai nei Radek Čech (2011) Čekijos prezidentų retorikos tyrime, šiame tyrime šie indikatoriai leido atskleisti, kad JAV prezidentų inauguracinių kalbų retoriką daugiausiai lėmė prezidentų individualus stilius, nors tam tikros visuomenės realijos (pavyzdžiui, karas ar ekonominė recesija) taip pat turėjo įtakos.

³⁸ Indikatorius λ yra teksto dažninės struktūros įvertis (Čech, 2011). Kartais naudojamas kaip žodyno turtingumo matas (Fang ir Liu, 2015).

³⁹ Žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu ir Altmann, 2015). Tekstų dažninės struktūros skirtumai ir panašumai leidžia tirti individualųjį stilių, žanrus, įvairias kalbos atmainas ir panašiai.

⁴⁰ Teminės koncentracijos variantas (plačiau žr. 1.4.2.5 sk.). Teksto teminė koncentracija vertinama per retesnių (pagrindinėms kalbos dalims priklausančių) žodžių, įsiterpusių tarp labai dažnų (paprastai – funkcinų) iki h-taško (žr. 1.4.2.1 sk.). Neretai pasitaiko, kad tarp labai dažnų žodžių (paprastai – funkcinų) nebūna įsiterpę pagrindinėms kalbos dalims priklausančių žodžių, tam teminės koncentracijos vertė būna lygi 0 (Kubát, Matlach ir Čech, 2014, 52). Antrinė teminė koncentracija leidžia išspręsti šią problemą.

⁴¹ Aktyvumas čia vertinamas veiksmažodžių ir būdvardžių santykiu tekste (plačiau žr. 1.4.3.2 sk.).

Paminėtinas ir Italijos prezidentų kalbų, pasakytų Naujųjų metų proga 1949–2012 m., retorikos tyrimas, kuriam taip pat naudotas aktyvumo indikatorius (Zörnig ir Altmann, 2016). Šis indikatorius leido išskirti prezidentus, kurie turėjo didelės įtakos Italijos visuomenėje ir kurių kalboms būdingas aiškumas (tai rodė didelės aktyvumo vertės, rodančios dinamišką kalbėjimą), bei prezidentus, kurių kalbose pasireiškė gyvumo stoka, susijusi, anot autorių, su nusivylimais karjeros srityje. Teminės koncentracijos indikatorius⁴² taip pat naudotas retorikos tyrimams, kai koncentruojamasi ties ideologijos raiška (Chen ir Liu, 2015). Šis indikatorius parodė, jog aukščiausios ir žemiausios teminės koncentracijos vertės priklauso nesutinkančioms nuomonėms, o statistiškai reikšmingas skirtumas teminės koncentracijos verčių kontekste rastas tarp daugumos nuomonės ir nesutinkančios nuomonės. Be to, kadangi statistiškai reikšmingų skirtumų tarp daugumos nuomonės raiškos nenustatyta, autorių nuomone, tai rodo, kad teminės koncentracijos vertės šiame tyrime gali žymėti ideologines pozicijas.

Be retorikos tyrimų, kiekybiniai indikatoriai naudojami ir automatiniam tekstų grupavimui į žanrus, tipus ar kitas kategorijas. Teminės koncentracijos indikatorius sėkmingai leido automatiškai sugrupuoti anglų ir kinų kalbų tekstus pagal, kaip autoriai įvardija, tekstų tipus – spaudą, oficialiuosius dokumentus ir grožinę literatūrą (Chen ir Liu, 2018). Iš esmės tai būtų galima laikyti FS (spauda iš dalies atitiktų publicistinį FS, oficialieji dokumentai – administracinį FS ir grožinė literatūra – meninį FS). Šis tyrimas atskleidė, kad teminės koncentracijos indikatorius leidžia atskirti tekstus pagal teminį intensyvumą, kuris yra nevienodai stipriai išreikštas skirtingų FS tekstuose. Šio tyrimo rezultatais buvo pasiremta įtraukiant teminės koncentracijos indikatorių lietuvių kalbos FS kalbinės raiškos požymių kiekybinę analizę.

Kaip jau buvo minėta, neretai tyrimuose naudojami ne pavieniai indikatoriai, o įvairaus dydžio jų rinkinys. Indikatorių rinkinys (MATTR (angl. *Moving Average Type-Token Ratio*)⁴³, MWTTRD (angl. *Moving Window Type-Token Ratio Distribution*)⁴⁴, teminė koncentracija⁴⁵, antrinė

⁴² Žymi, kaip tekstas yra sukoncentruotas į temą (Čech, Garabík ir Altmann, 2015; Kubát, 2015, 57–58). Plačiau žr. 1.4.2.5 sk.

⁴³ Nuo teksto ilgio nepriklausomas TTR variantas (plačiau žr. 1.4.2.4 sk.).

⁴⁴ Dar vienas TTR variantas. Leidžia tirti TTR verčių pasiskirstymą nustatyto dydžio žodžių segmentams (Kubát ir Milička, 2013).

⁴⁵ Žymi, kaip tekstas yra sukoncentruotas į temą (Čech, Garabík ir Altmann, 2015; Kubát, 2015, 57–58). Plačiau žr. 1.4.2.5 sk.

teminė koncentracija⁴⁶, proporcinė teminė koncentracija⁴⁷, tarpveiksmažodiniai atstumai⁴⁸, vidutinis žodžio ilgis tekste⁴⁹, aktyvumo indikatorius⁵⁰ ir deskriptyvumo indikatorius⁵¹) naudotas ir čekų kalbos tekstų grupavimui pagal žanrus Mirosław Kubát (2015). Šio tyrimo rezultatai parodė, kad kiekvienas iš indikatorių atskirai geba iš dalies skirti tirtus tekstus, bet geriausiai šiuos tekstus pagal žanrą leido atskirti aktyvumas, tarpveiksmažodiniai atstumai ir vidutinis žodžio ilgis tekste (Kubát, 2015, 118). Kitaip tariant, šių 3 indikatorių raiška įvairių žanrų tekstuose čekų kalba skyrėsi labiausiai. Šiuo tyrimu irgi remtasi pasirenkant indikatorius lietuvių kalbos FS kalbinės raiškos kiekybinei požymių analizei. Panašus į aptartą tyrimas atliktas ir skirtingų tipų⁵² tekstų persų kalba (eilėraščiai, laikraščio straipsniai, religiniai tekstai, apsakymai) grupavimui, tik naudotas kitoks indikatorių rinkinys: TTR, žodyno tankumo⁵³, R_1 ⁵⁴, h-taško⁵⁵ ir lambda⁵⁶ indikatoriai (Miangah ir Rezai, 2016). Šio tyrimo rezultatai parodė, kad TTR ir lambda indikatoriai galėtų tikti persų kalbos tekstų žanrų skyrimui (Miangah ir Rezai, 2016). Dar kitoks indikatorių rinkinys (santykinis RR⁵⁷, a ir R_1) naudotas automatiniam tekstų serbų kalba grupavimui, vėlgi, anot

⁴⁶ Teminės koncentracijos variantas, plačiau žr. 33-ą išnašą.

⁴⁷ Alternatyva teminei koncentracijai ir antrinei teminei koncentracijai. Indikatoriaus privalumas – galimybė palyginti tekstus, turinčius tik po 1 pagrindinės kalbos dalims priklausantį žodį, įsiterpusį tarp labai dažnų (paprastai – funkcinių) žodžių naudojant statistinio reikšmingumo testą (Kubát, 2015, 63).

⁴⁸ Vertina vidutinį atstumą tekste tarp 2 iš eilės einančių veiksmažodžių arba kalbos ar teksto tyrimų kontekste, teksto sintaksinės struktūros kompleksiskumą (plačiau žr. 1.4.3.3 sk.).

⁴⁹ Teksto sudėtingumo (suprantamumo prasme) įvertis (angl. *readability*); plačiau žr. 1.4.3.1 sk.)

⁵⁰ Vertinamas veiksmažodžių ir būdvardžių santykiu tekste, plačiau žr. 1.4.3.2 sk.).

⁵¹ Indikatorius priešingas aktyvumui, t. y. vertinamas būdvardžių ir veiksmažodžių santykiu tekste (plačiau žr. 1.4.3.3 sk.).

⁵² Iš esmės tyrimui naudotas tekstynas – žanrų ir funkcinių stilių mišinys. Galima daryti prielaidą, kad tyrimo autoriai tiesiog bandė surinkti kaip galima skirtingesnius tekstus.

⁵³ Vertinamas daiktavardžių ir veiksmažodžių santykiu (Miangah ir Rezai, 2016).

⁵⁴ Žodyno turtingumo / įvairovės įvertis, plačiau žr. 1.4.2.2 sk.

⁵⁵ Žymi apytikslę ribą tarp labai dažnų (paprastai – funkcinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų dažninėje lentelėje (plačiau žr. 1.4.2.1 sk.).

⁵⁶ Teksto dažninės struktūros indikatorius. Teksto dažninė struktūra – žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu ir Altmann, 2015). Tad indikatorius lambda yra šio modelio įvertis.

⁵⁷ Žodyno koncentruotumo įvertis (plačiau žr. 1.4.2.5 sk.).

autorių, pagal tekstų tipus⁵⁸ (Zörnig, Kelih, ir Fuks, 2016). Pasirinkti indikatoriai parodė tyrime naudotų tekstų kalbinės raiškos įvairovę, kuria remiantis šiuos tekstus galima automatiškai atskirti vienus nuo kitų. Pastebėtina, kad autoriai tiesiog stengėsi surinkti kuo skirtingesnius tekstus, nelabai kreipdami dėmesį į formalias tekstų tipologijas. Atsižvelgiant į tyrimo rezultatus, iš esmės galima įžvelgti žanrinę tekstų diferenciaciją, t. y. tyrimui pasirinkti indikatoriai leido tirtuosius tekstus automatiškai atskirti pagal žanrus.

Autorystės kiekybiniuose tyrimuose indikatoriai taip pat naudojami. Galima išskirti tyrimą, kuriame indikatorių rinkinys naudotas siekiant įvertinti, kaip autoriaus individualus stilius išlieka net jam vartojant skirtingas kalbas (Juola, Mikros ir Vinsick, 2018). Tyrimas atliktas naudojant tekstus ispanų-anglų ir graikų-anglų kalbomis, o indikatorių rinkinį sudarė: TTR, hapax legomena⁵⁹ procentas tekste, Yule K charakteristika⁶⁰, h-taškas⁶¹, entropija⁶², vidutinis žodžio ilgis tekste⁶³, Writer's View⁶⁴, R₁ (žodžių

⁵⁸ Iš esmės tai vėlgį tekstų, priklausančių skirtingiems žanrams, funkciniais stiliams ir postiliams, mišinys – poezija, roko (muzikos stilius) dainų tekstai, įvairių žanrų šnekamosios kalbos tekstai (kasdieniai pokalbiai, gydytojo-paciento pokalbiai, kiti šnekamosios serbų kalbos tekstai), proza, moksliniai ir žurnalistiniai tekstai. Straipsnio autoriai, galima daryti prielaidą, tiesiog stengėsi surinkti kuo skirtingesnius tekstus.

⁵⁹ Žodžiai, tekste pasikartojantys 1 kartą (Kubát, Matlach ir Čech, 2014, 20).

⁶⁰ Yule K įvertina leksinius pasikartojimus tekste (Miranda-García ir Calle-Martín, 2005).

⁶¹ Žymi tašką dažninėje lentelėje, kur žodžio ar žodžio formos dažnis yra lygus jo pozicijai toje lentelėje, o kalbos ir teksto tyrimų kontekste h-taškas žymi apytikslę ribą tarp labai dažnų (paprastai – funkciinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų (plačiau žr. 1.4.2.1 sk.).

⁶² Entropija lingvistikoje interpretuojama nevienodai. Žodžių pasikartojimų dažnumų tyrimuose entropija laikoma įvairovės, heterogeniškumo, neapibrėžtumo matu (Popescu, 2009, 173). Kuo entropijos vertė didesnė, tuo žodynas yra labiau sukoncentruotas į keletą žodžių. Kitaip tariant, didelė entropija žymi mažą žodyno turtingumą ir atvirkščiai. Ši entropijos interpretacija yra vyraujanti, tačiau esama ir kitos, kuri teigia, kad entropija žymi monotoniškumą arba stereotipiškumą (Altman ir Köhler, 2015, 36). Tai aiškinama tuo, kad jei entropijos vertė yra labai didelė, tai rodo, kad didžioji dalis žodžių pasikartojimų dažnumų yra susitelkę į kelis ar viso labo 1 žodį, kitų žodžių dažniai yra maži. Tad šioje pasikartojimų struktūroje įžvelgiama monotoniškumą arba stereotipiškumą.

⁶³ Lingvistiniuose tyrimuose ši charakteristika naudojama teksto sudėtingumo (suprantamumo prasme) įvertinimui (angl. *readability*; plačiau žr. 1.4.3.1 sk.)

⁶⁴ Šis indikatorius naudojamas balanso tarp labai dažnų (paprastai – funkciinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų tekste įvertinimui (Popescu, Mačutek ir Altmann, 2009, 26; Kubát, Matlach ir Čech, 2014, 44–45).

turtingumo arba įvairovės indikatorius, plačiau žr. 1.4.2.2 sk.), RR^{65} , RR_{mc} (žodyno koncentracijos įvertis, plačiau žr. 1.4.2.3 sk.), žodžių dažnių-rangų kreivės ilgis⁶⁶, žodžių dažnių-rangų kreivės ilgio indeksas R^{67} , Adjusted Modulus⁶⁸, Gini koeficientas⁶⁹, perteklumo (angl. *redundancy*) charakteristika⁷⁰. Šio tyrimo rezultatai parodė, jog autoriaus stilius išlieka net jam ir vieną kalbą keičiant kita, o tarp indikatorių verčių suporuotoms kalboms (ispanų-anglų ir graikų-anglų) nustatytas statistiškai reikšmingas ryšys. Tad kiekybiniai indikatoriai leido sukurti prielaidas autorystės priskyrimui, kai autorius rašo ne viena kalba (Juola, Mikros ir Vinsick, 2018). Didžioji dalis šiame tyrime naudotų indikatorių nebuvo pasirinkti disertacijos tyrimui, kadangi, nors matematinės jų išraiškos ir yra aiškios, tačiau interpretacijos kalbos atžvilgiu nėra pakankamai ištirtos.

Be retorikos analizės, tekstų grupavimo ir autorystės tyrimų, kiekybiniai indikatoriai naudojami ir tekstų analizei. Tekstų analizės kontekste paminėtinas poezijos slovakų kalba tyrimas, kuriame tirta poezijos tekstų žodyno turtingumo raiška (Popescu, Čech ir Altmann, 2011a). Šiame tyrime taip pat naudotas indikatorių rinkinys: Gini koeficientas⁷¹, R_4^{72} , R_1 (žodyno turtingumo matas, plačiau žr. 1.4.2.2 sk.), santykinė entropija⁷³, RR_{mc} (žodyno

⁶⁵ RR (angl. *Repeat Rate*) yra žodyno koncentruotumo įvertis (Kubát, Matlach ir Čech, 2014, 16). Šis įvertis yra priešingas žodyno turtingumui. Kitaip sakant, jei tekstui būdinga didelė žodyno koncentracija, tai jam taip pat būdingas mažas žodyno turtingumo ir atvirkščiai (plačiau žr. 1.4.2.3 sk.)

⁶⁶ Žodyno turtingumas įvertis, paremtas dažnių-rangų (pozicijų) dažninėje lentelę pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, 35–36; Juola, Mikros ir Vinsick, 2018)

⁶⁷ Žodyno turtingumo įvertis, išvestas remiantis žodžių dažnių-rangų (pozicijų) dažninėje lentelėje pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, 36–37; Juola, Mikros ir Vinsick, 2018)

⁶⁸ Teksto dažninės struktūros (žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu ir Altmann, 2015) indikatorius, kuris, autoriaus teigimu, nepriklauso nuo teksto ilgio (Popescu, et al., 2010, p. 4–5).

⁶⁹ Matematiškai tai – statistinės dispersijos įvertis, o lingvistikoje paprastai naudojamas kaip vienas žodyno turtingumo matų (Kubát, Matlach ir Čech, 2014, 30–31; Juola, Mikros ir Vinsick, 2018).

⁷⁰ Kalbos našumo įvertis (Jaeger, 2010; Stilp, 2011).

⁷¹ Matematiškai tai – statistinės dispersijos įvertis, o lingvistikoje paprastai naudojamas kaip vienas žodyno turtingumo matų (Kubát, Matlach ir Čech, 2014, 30–31; Juola, Mikros ir Vinsick, 2018).

⁷² Žodyno turtingumo matas, atvirkščias Gini koeficientui (Kubát, Matlach ir Čech, 2014, 34).

⁷³ Normalizuota entropija. Žodžių pasikartojimų dažnumų tyrimuose entropija laikoma įvairovės, heterogeniškumo, neapibrėžtumo matu (Popescu, 2009, 173). Kuo entropijos vertė didesnė, tuo žodynas yra labiau sukcentruotas į keletą žodžių.

koncentracijos įvertis, plačiau žr. 1.4.2.3 sk.) ir žodžių dažnių-rangų kreivės ilgio indeksą R^{74} . Kaip galima matyti iš didelio indikatorių rinkinio, autoriai siekė poezijos tekstų žodyno turtingumo raišką tirti įvairiais aspektais. Kitaip tariant, nors visi tyrime naudoti indikatoriai naudoti vienos savybės – žodyno turtingumo – analizei, tačiau visų šių indikatorių matematinės formulės yra skirtingos, tad ir rezultatų interpretacijos kalbos požiūriu nėra identiškos. Anot straipsnio autorių, norint įvertinti teksto žodyno turtingumą, svarbu pasitelkti bent keletą tam tikslui skirtų indikatorių, kadangi jie kreipia dėmesį į skirtingus žodyno turtingumo aspektus (Popescu, Čech ir Altmann, 2011a). Tad, remiantis šiuo tyrimu, disertacijoje pristatomame lietuvių kalbos FS kalbinės raiškos tyrimui pasirinktas ne vienas indikatorius, o jų rinkinys, kad būtų galima iširti skirtingus FS kalbinės raiškos aspektus.

Panašų į aptartąjį tyrimą grupė mokslininkų atliko tekstams italų kalba (Tuzzi, Popescu ir Altmann, 2010). Šiame tyrime be jau žinomų indikatorių (R_1 (žodyno turtingumo matas, plačiau žr. 1.4.2.2 sk.), *Writer's View*⁷⁵, su žodžių dažnių-rangų kreivės ilgiu⁷⁶ susiję indikatoriai ir pan.) taikyti ir keletas retesnių – santykinės asociatyviosios koncentracijos įvertis⁷⁷, santykinio susietumo įvertis⁷⁸ bei kiti. Pasirinkti indikatoriai leido tekstus iširti keletu aspektų – žodyno koncentracijos, asociacijos tarp žodžių ir kitų kalbos vienetų, balanso tarp labai dažnų tiriamųjų kalbos vienetų (pavyzdžiui, žodžių) ir retesnių, tekstų dažninės struktūros (žodžių ir kalbos dalių

Kitaip tariant, didelė entropija žymi mažą žodyno turtingumą ir atvirksčiai. Ši entropijos interpretacija yra vyraujanti, tačiau esama ir kitos, kuri teigia, kad entropija žymi monotoniškumą arba stereotipiškumą (Altman ir Köhler, 2015, 36). Šiame tyrime entropija naudojama žodyno turtingumui matuoti.

⁷⁴ Žodyno turtingumas įvertis, paremtas dažnių-rangų (pozicijų) dažninėje lentelę pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, 35–36; Juola, Mikros ir Vinsick, 2018).

⁷⁵ Balanso tarp labai dažnų (paprastai – funkcinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų tekste įvertis (Popescu, Mačutek ir Altmann, 2009, 26; Kubát, Matlach ir Čech, 2014, 44–45).

⁷⁶ Žodyno turtingumas įvertis, paremtas dažnių-rangų (pozicijų) dažninėje lentelę pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, p. 35–36; Juola, Mikros ir Vinsick, 2018).

⁷⁷ Paremtas sutaptimis (angl. *co-occurrences*) tam tikrame kontekste. Šiame tyrime naudojamos lemų sutaptys, o kontekstas – sakinys. Iš sutapčių sukonstravus tinklą, kai viršūnes („taškus“) vaizduoja žodžiai, o ryšius („taškus“ jungiančios linijos) – žodžių greta sakinyje pasikartojimai, tai santykinės asociatyviosios koncentracijos įvertis vertina vidutinį ryšių skaičių iš greta pasikartojančių žodžių sudarytame tinkle (Tuzzi, Popescu ir Altmann, 2010).

⁷⁸ Šis įvertis rodo esamų komponentų (žodžių sutapčių grupių tinkle) ir visų galimų komponentų žodžių sutapčių tinkle santykį (Tuzzi, Popescu ir Altmann, 2010).

atžvilgiu). Šiuo tyrimu taip pat buvo pasiremta disertacijos tyrimui sudarant indikatorių rinkinį. Kadangi daugumos, ypač retesnių šiame tyrime naudotų indikatorių skaičiavimų interpretavimas kalbos požiūriu nėra iki galo ištirtas, į disertacijos tyrimui skirtą indikatorių rinkinį įtraukti tik tie indikatoriai, kurių interpretacija kalbos atžvilgiu yra pakankamai ištirta.

Paminėtinas ir leksinių požymių dinamikos anglų kalba rašytose daktaro disertacijose, priklausančioms skirtingoms mokslo sritims, tyrimas (Xiao ir Sun, 2018). Jame naudotą kiekybinių indikatorių rinkinį sudarė TTR, h-taškas⁷⁹, R_1 (žodyno turtingumo matas, plačiau žr. 1.4.2.2 sk.) ir Writer's View⁸⁰. Tyrimo rezultatai parodė, kad TTR ir R_1 žymi aiškius skirtumus tarp humanitarinių ir gamtos mokslų daktaro disertacijų: humanitarinių mokslų daktaro disertacijoms buvo būdinga didesnė leksikos įvairovė (didesnės TTR vertės), o gamtos mokslų daktaro disertacijoms – didesnė turinio (retesnių, priklausančių pagrindinėms kalbos dalims) žodžių gausa (didesnės R_1 vertės) (Xiao ir Sun, 2018). Kadangi kalbinės raiškos požymiai, kuriuos žymi minėti indikatoriai buvo aktualūs disertacijos kontekste, jie įtraukti į tyrimui naudotų indikatorių rinkinį (tik vietoj TTR pasirinktas nuo teksto ilgio nepriklausomas MATTR variantas). Šiek tiek netradiciniame tyrime žodyno turtingumo (taikyta indikatoriaus R_4 ⁸¹ modifikacija) bei teminės koncentracijos kiekybiniai indikatoriai taikyti fetišistinės literatūros anglų kalba analizei (Wilson, 2009). Tyrimo rezultatai parodė, kad fetišistinei literatūrai nebūtinai būdingas mažesnis žodyno turtingumo nei klasikinei anglų grožinei literatūrai. Be to, teminės koncentracijos indikatorius parodė, kad fetišistinei literatūrai nėra būdingas koncentravimasis ties nedideliu rinkiniu raktinių žodžių.

Dar vienas mažiau įprastas tyrimas, kuriame naudoti kiekybiniai indikatoriai, – modernaus romano kinų kalba žanrų plėtotės tyrimas, kuriame

⁷⁹ Žymi apytikslę ribą tarp labai dažnų (paprastai – funkcinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų dažninėje lentelėje (plačiau žr. 1.4.2.1 sk.).

⁸⁰ Balanso tarp labai dažnų (paprastai – funkcinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų tekste įvertis (Popescu, Mačutek ir Altmann, 2009, 26; Kubát, Matlach ir Čech, 2014, 44–45).

⁸¹ Žodyno turtingumo matas, atvirkščias Gini koeficientui (Kubát, Matlach ir Čech, 2014, 34). Gini koeficientas lingvistikoje paprastai naudojamas kaip vienas žodyno turtingumo matų (Kubát, Matlach ir Čech, 2014, 30–31; Juola, Mikros ir Vinsick, 2018).

naudoti indikatoriai λ ⁸² ir a ⁸³ pakankamai gerai gebėjo aptikti žanrų pokyčius (Zhang ir Liu, 2015). Netradicinių taikomųjų tyrimų kontekste paminėtinas ir tyrimas, kuriame buvo tikrinama hipotezė, kad teksto vertimo žodyno turtingumas yra mažesnis nei teksto originalo (Fang ir Liu, 2015). Tyrime naudoti tekstai anglų ir kinų kalbomis, taikyti indikatoriai STTR (angl. *standardized type-token ratio*)⁸⁴ ir λ ⁸⁵. Pagal STTR gauti rezultatai parodė, jog žodyno turtingumas verstuose į anglų kalbą grožinės literatūros tekstuose yra mažesnis nei neverstiniuose grožiniuose tekstuose, nesvarbu, ar vertėjo gimtoji kalba yra anglų, ar ne.

Paskutinė apžvelgiamų taikomųjų tyrimų grupė – tipologiniai tyrimai. Viename jų žodžių pasikartojimų dažnumų 20-ye kalbų ypatumams tirti naudoti indikatoriai h-taškas⁸⁶, žodžių dažnių-rangų kreivės ilgis⁸⁷, hapax legomena⁸⁸ paremti indikatoriai, entropija, ir RR⁸⁹ (Popescu, Mačutek ir Altmann, 2009). Šie indikatoriai leido atlikti apžvalginę žodyno turtingumo raiškos analizę, įvertinti tyrime naudotas kalbas pagal jų analitiškumą-sintetiškumą, atlikti pirminį kalbos lygmenų (kategorijų) ir žodžių dažnių įvairovės bei tipologijos tyrimus, įvertinti tekstus pagal jų nominalumo raišką. Kitame panašiam tyrime buvo analizuojami poezijos tekstų žodžių dažnių-rangų (pozicijų) dažninėje lentelėje pasiskirstymai, kuriems naudoti indikatoriai apėmė h-tašką, indikatorius a ir b ⁹⁰, taip pat RR ir RR_{mc} (Popescu,

⁸² Teksto dažninės struktūros įvertis (Čech, 2011). Kartais naudojamas kaip žodyno turtingumo matas (Fang ir Liu, 2015). Teksto dažninė struktūra – žodžių ir žodžių formų dažnių ir pozicijų (rangų) dažninėje lentelėje modelis (Čech, 2015; Popescu ir Altmann, 2015).

⁸³ Nuo teksto ilgio nepriklausomas h-taško variantas (Popescu et al., 2017).

⁸⁴ Standartizuotas iteracijos indeksas, t. y. įprastas TTR skaičiuojamas kas 1000-ajam žodžiui (Fang ir Liu, 2015). Šis TTR variantas, laikoma, sumažina TTR priklausomybę nuo teksto ilgio. STTR realizuotas įrankyje WordSmith Tools (<https://lexically.net>).

⁸⁵ Indikatorius λ yra teksto dažninės struktūros (žodžių dažnių ir pozicijų pasiskirstymo dažninėje lentelėje modelis) įvertis (Čech, 2011). Kartais naudojamas kaip žodyno turtingumo matas (Fang ir Liu, 2015).

⁸⁶ Žymi apytikslę ribą tarp labai dažnų (paprastai – funkcinių) ir retesnių (paprastai – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų dažninėje lentelėje (plačiau žr. 1.4.2.1 sk.).

⁸⁷ Žodyno turtingumas įvertis, paremtas dažnių-rangų (pozicijų) dažninėje lentelėje pasiskirstymo kreive (Kubát, Matlach ir Čech, 2014, 35–36; Juola, Mikros ir Vinsick, 2018).

⁸⁸ Žodžiai, tekste pasikartojantys 1 kartą (Kubát, Matlach ir Čech, 2014, 20).

⁸⁹ Žodyno koncentruotumo įvertis (plačiau žr. 1.4.2.3 sk.).

⁹⁰ Indikatoriai a ir b yra nuo teksto ilgio nepriklausomi h-taško variantai (Popescu et al., 2017). Apie h-tašką plačiau žr. 1.4.2.1 sk.

et al., 2017). Be to, h-taškas ir jo modifikacija a^{91} taikyti tipologinių *zhuang*⁹² kalbos požymių tyrime (Wei ir Liu, 2018). Šis tyrimas atskleidė, kad pagal žodžių dažnių pasiskirstymą *zhuang* kalbos yra artimos polineziečių kalbų šeimai.

Šiame poskyryje pristatyti aktualūs disertacijos kontekste taikomieji tyrimai – pagrindas, kuriuo buvo remiamasi lietuvių kalbos FS kalbinės raiškos kiekybiniam tyrimui pasirenkant kiekybinius indikatorius. Remiantis apžvelgtais darbais, disertacijos tyrimui pasirinkti indikatoriai, nepriklausantys nuo teksto ilgio (arba kurių priklausomybė nuo teksto ilgio nėra didelė), žymintys skirtingas teksto/kalbos savybes bei kurių interpretacija kalbos požiūriu yra pakankamai ištirta. Kadangi indikatoriai atkreipia dėmesį į skirtingus kalbinės raiškos, pavyzdžiui, žodyno turtingumo, aspektus, lietuvių kalbos FS kalbinės raiškos požymių kiekybiniam tyrimui pasirinktas ne vienas indikatorius, o jų rinkinys. Tad pasirinkus apžvelgtų tyrimų rezultatais bei publikacijose pateiktomis rekomendacijomis, lietuvių kalbos FS kalbinės raiškos požymių kiekybiniam tyrimui pasirinkti 9 indikatoriai: a , R_1 , RR_{mc} , MATTR, teminė koncentracija (TK), vidutinis žodžių arba žodžių formų ilgis (VŽFI), aktyvumas (A), deskriptyvumas (D) ir tarpveiksmažodiniai atstumai (VA). Plačiau apie šiuos indikatorius žr. 1.4.2-1.4.3. sk.

⁹¹ Žr. 90-ą išnašą.

⁹² Viena kalbų grupių, kurios kalbomis kalbama Kinijoje (Wei ir Liu, 2018).

2. TYRIMO MEDŽIAGA

2.1 Tiriamosios medžiagos pagrindas

Šioje disertacijoje buvo apsispręsta tirti 3 FS kalbinę raišką – administracinio, mokslinio ir publicistinio, neatsižvelgiant į postilius ir žanrus. Tad tyrimui atlikti buvo parengti 3 tiksliniai tekstynai: administracinio FS tekstynas (AST), mokslinio FS tekstynas (MST) ir publicistinio FS tekstynas (PST). Administracinio stiliaus ir mokslinio stiliaus tekstynų pagrindas – Dabartinės lietuvių kalbos tekstynas⁹³ (1-oji versija, ~140 mln. žodžių; toliau – DLKT), sudarytas VDU Kompiuterinės lingvistikos centro darbuotojų. Publicistinio stiliaus tekstyno pagrindas – projekto „Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas (PASTOVU)“⁹⁴ (nr. LIP-027/2016) metu sukurtas 2014–2016 m. delfi.lt tekstynas.

DLKT pagrindu pasirinktas dėl to, kad tai didelis, bendro pobūdžio tekstynas, kuriam būdinga tekstų įvairovė, apimanti įvairių stilių tekstus. DLKT sudarytojai laikėsi nuomonės, kad geresnis yra didesnis tekstynas, nes taip didėja tikimybė užfiksuoti tiek tipišką, tiek rečiau vartojamų kalbinių konstrukčių vartoseną, kartu neatmetant tekstyno reprezentatyvumo (Kovalevskaitė, 2006). Didesniame tekстыne labiau tikėtina rasti didesnę tekstų tipų, žanrų ir temų įvairovę. Kita vertus, tekstyno reprezentatyvumui daug įtakos turi tekstyno šaltinių įvairovė, o ji priklauso nuo šaltinių atrankos kriterijų, kurių esama pačių įvairiausių. Tokie kriterijai gali būti elitiškumas, skaitomumas, demografiniai rodikliai, prieinamumas, tam tikri tekstyno sudarytojų sprendimai, pagrįsti kalbos vartojimo ypatumais ir pan. (Marcinkevičienė, 2000). DLKT sudarytas remiantis dydžio, reprezentatyvumo ir tęstinumo kriterijais (Kovalevskaitė, 2006). Tad DLKT administracinių tekstų dalis pasirinkta administracinio stiliaus tekstyno pagrindu, o negrožinės literatūros dalis, sudaryta iš mokomosios ir pažintinės (mokslo populiarinamosios) literatūros tekstų, pasirinkta mokslinio FS tekstyno pagrindu.

⁹³ Paieška tekстыne prieinama adresu <http://tekstynas.vdu.lt/tekstynas/>. Disertacijos autoriui buvo leista tyrimo tikslais naudotis šiuo tekstynu pateikus prašymą VDU Kompiuterinės lingvistikos centro vadovui dr. Andriui Utkai.

⁹⁴ Projekto tinklalapis pasiekiamas adresu <http://mwe.lt/>.

Publicistinio FS tekstyno pagrindu pasirinktas ne DLKT, bet projekto PASTOVU reikmėms sukauptas delfi.lt tekstynas. Pasirinkimas grindžiamas tuo, kad publicistinis FS, kurio tekstai fiksuoja pokyčius valstybės, visuomenės, kultūros ir kitose srityse, patys taip pat imlus pokyčiams (Bitinienė, 2006; Butkutė, 2008; Leonavičienė, 2010), o publicistiniai tekstai, įtraukti į DLKT, publikuoti tik iki 2008 m. imtinai. Projekto PASTOVU („Lietuvių kalbos pastoviųjų žodžių junginių automatinis atpažinimas“) tikslas – dabartinės rašytinės lietuvių kalbos pastoviųjų žodžių junginių tyrimo metodikos sukūrimas bei tekstynu paremto lietuvių kalbos kolokacijų žodyno parengimas (daugiau žr. <http://mwe.lt/>). Įgyvendinant projektą buvo sukauptas tekstynas, sudarytas iš 2014–2016 m. naujienų portale Delfi publikuotų straipsnių. Ši delfi.lt tekstyną sudaro ~72 mln. žodžių, tekstai surinkti iš 11 naujienų portalo Delfi skilčių: DELFI veidai, projektai, DELFI mokslas, DELFI auto, sportas, DELFI gyvenimas, DELFI FIT, DELFI pilietis, verslas, naujausios ir kita (tekstai, kurie buvo nepriskirti nei vienai išvardintai kategorijai) (Bumbulienė, Mandravickaitė, Boizou ir Krilavičius, 2017). Tad šiam tekstynui būdinga tiek teminė, tiek žanrinė įvairovė. Kadangi disertacijoje tiriama kalbinė raiška ne tam tikro žanro, o publicistinio FS apskritai, delfi.lt tekstynas yra tinkamas kaip pagrindas tiksliniam publicistinio stiliaus tekstynui.

2.2 Tiriamosios medžiagos parengimo žingsniai

FS tikslinių tekstynų disertacijos tyrimui parengimas vyko keliais žingsniais. Kadangi tekstynų pagrindai skyrėsi, tekstų parengimo žingsniai atskiriems tekstynams taip pat buvo nevienodi. Reikia paminėti, kad bazinis šioje disertacijoje pasirinktais metodais atliktos analizės vienetas yra tekstas, tad disertacijos autorės sudarant tikslinius tekstynus labiau atsižvelgta į tekstynų subalansuotumą tekstų skaičiaus atžvilgiu nei žodžių skaičiaus atžvilgiu. Kitaip tariant, sudarant tikslinius tekstynus (AST, MST ir PST), iš esmės nekreipta dėmesio į atskirą tekstyną sudarančių žodžių skaičių. Labiau žiūrėta, kad tekstynai būtų pakankamai dideli tekstynus sudarančių tekstų skaičiaus atžvilgiu.

Taigi tyrimui sudarytų tikslinių tekstynų parengimo žingsniai buvo šie:

1. Tekstų atranka pagal disertacijos autorės nusistatytus kriterijus (kiekvieno FS tekstynui taikyti tekstų atrankos kriterijai plačiau aprašomi 2.3, 2.4, ir 2.5 sk.).

2. Sukaupti tekstai buvo pateikti .xml (tekstai, paimti iš DLKT), .pdf (daktaro disertacijų santraukos, įtrauktos į tikslinį MST) arba .txt (tekstai, paimti iš delfi.lt tekstyno) formatais. Tyrimo atlikimui reikėjo, kad tekstai būtų pateikti paprasto tekstinio failo formatu (.txt), tad .xml ir .pdf tekstai buvo konvertuoti į paprastus tekstinius failus, naudojant įrankį Calibre⁹⁵.
3. Sukauptuose tikslinių dokumentų failuose, paimtuose iš DLKT, kartu su tekstais buvo pateikti ir metaduomenys, t. y. informacija apie tekstą, pavyzdžiui, teksto autorius, publikavimo metai, žanras ir panašiai. Dalis tekstų taip pat buvo pateikti kartu su tam tikra papildoma informacija, pavyzdžiui, turiniu, indeksu, lentelėmis ir pan. Ši informacija kartu su metaduomenimis, pateiktais faile kartu su tekstu, bei įvairūs ne raides ir skaičius žymintys (angl. non-alphanumeric) simboliai buvo pašalinti, kad neiškraipytų tekstų kalbinę raišką žyminčių indikatorių skaičiavimo rezultatu, t. y. kad nebūtų laikoma tiriamąja medžiaga. Tai šios disertacijos autorės buvo atlikta iš dalies automatiškai (naudojant statistinį paketą R ir reguliariųjų išraiškų taisykles (Karttunen, Chanod, Grefenstette ir Schille, 1996)), iš dalies rankiniu būdu.
4. Sukauptuose tekstuose buvo rasta ortografinių skirtumų, pavyzdžiui, išretintai parašytų žodžių, tekstą sudarančių simbolių kodavimo skirtingomis koduotėmis (Korpela, 2001) ir kita. Tad sukaupti tekstai tiksliniams tekstynams šios disertacijos autorės buvo standartizuoti. Išretintas tekstas pakeistas įprastu formatu (tai atlikta iš dalies automatiškai naudojant statistinį paketą R ir reguliariąsias išraiškas, iš dalies rankiniu būdu), visų tekstų simbolių koduotės suvienodintos, t. y. visi tekstus sudarantys simboliai konvertuoti į UTF-8 koduotę (Pike ir Thompson, 1993). Tai atlikta automatiškai, naudojant statistinį paketą R.
5. Kaip jau minėta, dviejų tyrimui sudarytų tikslinių tekstynų pagrindas – DLKT. Jo specifika yra ta, kad viename faile neretai yra daugiau nei vienas tekstas. Kadangi šioje disertacijoje atliekamo tyrimo bazinis tiriamasis vienetas yra būtent tekstas, failai, kuriuose buvo keletas tekstų, buvo suskaidyti taip, kad viename faile būtų tik vienas tekstas. Tai buvo atlikta automatiškai, naudojant statistinį paketą R.
6. Tekstų ilgio žodžiais patikra. Teminės koncentracijos (žr. 3.1.2.5 sk.), vieno iš tyrime naudojamų kiekybinių indikatorių, vertės, esant

⁹⁵ Įrankį galima parsisiųsti adresu <https://calibre-ebook.com/>.

trumpiems tekstams, yra labai didelės, t. y. priklausomos nuo teksto ilgio (Kubát ir Čech, 2016b). Tad į sudaromus tikslinius tekstynus įtraukti tik tie tekstai, kuriuos sudaro daugiau nei 200 žodžių⁹⁶.

7. Tekstus ruošiant kiekybinei analizei, reikėjo rasti ir pritaikyti teksto skaidymo į žodžius algoritmą (t. y. tekstus suskaidyti taip, kad vienoje eilutėje būtų vienas žodis), tinkamai skaidantį lietuviškas santrumpas, pavyzdžiui, vardo raidę ir po jos esantį tašką traktuojant kaip prasminį vienetą, t. y. taško neatskiriant ir nelaikant atskiru skyrybos ženklų ir pan. Tekstų skaidymas į žodžius atliktas automatiškai, naudojant statistinio paketo R biblioteką *quanteda* (Benoit et al., 2018).
8. Trijų indikatorių (aktyvumo ir deskriptyvumo bei tarpveiksmažodinių atstumų) skaičiavimui buvo būtina morfologinė informacija, tad tekstai buvo suanotuoti morfologiškai. Anotavimui naudotas projekto PASTOVU metu lietuvių kalbai pritaikytas morfologinio anotavimo įrankis ir to paties projekto metu sukurtas lietuvių kalbos morfologinio anotavimo modelis⁹⁷. Anotavimui taikytas Universaliųjų priklausomybių projekte⁹⁸ naudojamas morfologinių pažymų rinkinys:

- ADJ: būdvardis;
- ADP: prielinksnis;
- ADV:rieveiksmis;
- CCONJ: jungtukas (koordinavimas; angl. *coordinating conjunction*);
- INTJ: jaustukas;
- NOUN: daiktavardis;
- NUM: skaitmuo;
- PART: dalelytė;
- PRON: įvardis;
- PROP: tikrinis daiktavardis;
- PUNCT: skyrybos ženklas;
- SCONJ: jungtukas (šalutinė jungtis, angl. *subordinating conjunction*);
- SYM: simbolis;
- VERB: veiksmažodis;

⁹⁶ Plačiau žr. 1.4.2.5 sk.

⁹⁷ Disertacijos autorė dirbo šiame projekte jaunesniąja mokslo darbuotoja, turėjo prieigą prie šio viešai neteikiamo morfologinio anotatoriaus.

⁹⁸ Projekto informacija pasiekama adresu <http://universaldependencies.org/>.

- X: kita (paprastai šiuo ženkle žymimi neatpažinti žodžiai).

11-oje lentelėje pateikiama bazinė kiekvieno tikslinio tekstyno galutinių versijų statistinė informacija. Kaip galima matyti, visą tiriamąją medžiagą sudaro 19 002 tekstai (~ 36,4 mln. žodžių). Tekstynus sudaro skirtingas skaičius žodžių bei tekstų, tad gali kilti tyrimo patikimumo klausimas. Pagal matematinį didžiųjų skaičių dėsnį, didelių imčių vidurkiai priartėja visos populiacijos vidurkiui (Wasserman, 2004, 76). Tai reiškia, kad, surinkus pakankamą kiekį kiekvieną tiriamą FS atitinkančių tekstų, kalbinės raiškos tyrimo rezultatai bus artimi visų šiems funkciniais stiliams priklausančių tekstų vidurkiams. Kadangi lingvistiniai duomenys dažnai yra nutolę nuo normaliosios kreivės (Levshina, 2015, 18), kitaip sakant – yra asimetriški, gausūs itin didelių ar itin mažų verčių, išsiskiriančių iš bendro duomenų konteksto, tokiais atvejais rekomenduojama, kad imties dydis būtų didesnis nei 300 (Wilcox, 2012, 104). Kiekvieną tikslinį tekstyną sudaro po daugiau nei 300 tekstų, tad tyrimo rezultatai turėtų reprezentatyviai vaizduoti tiriamų FS kalbinę raišką. Skirtingas žodžių skaičius rezultatų irgi neturėtų iškreipti rezultatų, kadangi tyrime naudojamų kalbinę raišką žyminčių kiekybinių indikatorių vertėms teksto ilgis įtakos neturi arba beveik neturi (plačiau žr. 1.4 sk.).

11 lentelė. Tikslinių tekstynų charakteristikos⁹⁹

Tekstynas	Apimtis (žodžių sk.)	Apimtis (tekstų sk.)
Administracinio funkcinio stiliaus tekstynas	~ 5,8 mln.	4 527
Mokslinio funkcinio stiliaus tekstynas	~ 20,2 mln.	1 025
Publicistinio funkcinio stiliaus tekstynas	~ 10, 4 mln.	13 450
Iš viso	~ 36,4 mln.	19 002

Kadangi rengiant konkrečius tekstynus tyrimui buvo susidurta su skirtingo pobūdžio problemomis ir jiems parengti taikyti nevienodi žingsniai, tolesniuose skyreliuose aptariamas kiekvienas disertacijos tyrimui parengtas tikslinis tekstynas atskirai.

⁹⁹ Kadangi šitam tyrimui atlikti daugiau bazinių charakteristikų nei pateikta 11 lentelėje nebuvo reikalinga, tokios charakteristikos unikalių žodžių skaičius ar raidžių skaičius ir pan. nėra nepateiktos.

2.3 Tikslinių tekstynų sandara

2.3.1 Administracinio FS tekstynas

DLKT administracinės literatūros dalį, kurią buvo planuota imti šioje disertacijoje aprašomo administracinio FS kalbinės raiškos tyrimo pagrindu, sudarė tokie tekstai: Lietuvos Respublikos valstybiniai dokumentai (žanrai: deklaracija, aktas, ataskaita, dekretas, įstatymas, įsakymas, išvada, normatyvai, nutarimas, pranešimas ir pan.), Europos Sąjungos dokumentai (žanrai: aktas, ataskaita, chartija, dekretas, direktyva, išaiškinimas, išvada, konvencija, kreipimasis, programa ir pan.), bažnytiniai dokumentai (žanrai: bulė, chartija, kalba, kreipimasis, memorandumas, pareiškimas, sprendimas, sveikinimas, enciklika, laiškas ir pan.) ir dokumentai, priskirti kategorijai „Neapibrėžta“ (ją sudarė Lietuvos Respublikos Seimo posėdžių stenogramos). Tačiau šios disertacijos tyrimo tikslams DLKT administracinės literatūros dalimi nebuvo galima tiesiog pasinaudoti, nes ne visi šios dalies tekstai ir jų pateikimo formatai tiko disertacijoje iškeltų tikslų įgyvendinimui. Taigi disertacijos autorė koregavo tiek tekstų pateikimo formatą, tiek į tikslinį įtraukiamų tekstų kiekį.

Sudarant tikslinį administracinio FS tekstyną AST, visų pirma atlikta tekstų atranka. Į tiriamąjį tekstyną nuspręsta neįtraukti Lietuvos Respublikos Seimo posėdžių stenogramų (sąlytinė kalbos atmaina, be to, koreguota) ir bažnytinių dokumentų. Pastarųjų nuspręsta neįtraukti dėl dokumentų specifikos. Tad AST buvo sudarytas iš DLKT administracinių tekstų dalies – Lietuvos Respublikos valstybinių dokumentų ir Europos Sąjungos dokumentų. AST parengimui buvo taikyti visi 2.2 skyriuje pateikti žingsniai. Atlikus šiuos žingsnius, tikslinį AST sudarė 4527 tekstai (~ 5,8 mln. žodžių).

2.3.2 Mokslinio FS tekstynas

DLKT dalį, kuri buvo planuota imti šioje disertacijoje aprašomo mokslinio FS kalbinės raiškos tyrimo pagrindu, sudarė tokie tekstai: mokomoji literatūra (vadovėliai, metodinės ir kitos mokymo priemonės), mokslo populiarinamoji ir pažintinė literatūra (enciklopedijos, žinynai ir kt.) literatūra. Kadangi DLKT negrožinės literatūros dalyje (mokomieji, mokslo populiarinamieji ir pažintiniai tekstai DLKT buvo priskirti negrožinei literatūrai; publicistika čia laikoma atskira kategorija) nebuvo teorinio (mokslinio siaurąją prasme)

postilio tekstų, šioje disertacijoje aprašomo tyrimo tikslais tikslinis mokslinio FS tekstynas, greta DLKT tekstų, šios disertacijos autorės papildytas daktaro disertacijų, apgintų 2008–2018 m., santraukomis. Santraukos rankiniu būdu susirinktos iš Lietuvos akademinės elektroninės bibliotekos portalo¹⁰⁰. Į tikslinį mokslinio FS tekstyną MST įtrauktos daktaro disertacijų santraukos iš visų mokslo sričių. Visos santraukos tiksliniam MST atsirinktos pagal šiuos kriterijus:

1. **Laikotarpio kriterijus:** pasirinktos santraukos disertacijų, apgintų per paskutinius 10 metų, t. y. 2008–2018 metais.
2. **Kalbos, kuria parašyta santrauka, kriterijus:** pasirinktos santraukos, parašytos lietuvių kalba (praktiškai apie pusę atvejų viename dokumente/faile buvo pateiktos 2 santraukos – lietuvių ir anglų kalbomis; pagrindinis skirtumas tarp santraukų, parašytų dviem kalbomis – teksto lietuvių ir anglų kalba proporcijos, t. y. vienais atvejais teksto apimties atžvilgiu lietuviškas ir angliškas tekstai yra maždaug vienodos apimties, kitais atvejais santraukos viena kuria kalba (anglų arba lietuvių) apimtis yra daug didesnė nei rašytos kita kalba). Į tikslinį mokslinio FS tekstyną įtrauktos tik lietuvių kalba parašytos santraukos. Jei tame pačiame faile pateikta ir santrauka anglų kalba, tekstyno paruošimo analizei etape ji pašalinama rankiniu būdu.
3. **Santraukos prieinamumo kriterijus:** pasirinktos santraukos, turinčios atvirąją prieigą.

Pagal minėtus kriterijus surinkta 617 daktaro disertacijų santraukų: 68 – medicinos ir sveikatos mokslų, 199 – gamtos mokslų, 75 – humanitarinių mokslų, 164 – socialinių mokslų, 90 – technologijos mokslų ir 21 – žemės ūkio mokslų.

Rengiant ir tvarkant tikslinį MST, taikyti 6 iš 7 žingsnių (netaikytas 5-asis), išvardytų 2.2 sk. Tad tikslinį mokslinio FS tekstyną sudaro 408 DLKT negrožinės literatūros dalies tekstai (mokomojo ir mokslo populiarinamojo postilių tekstai) ir 617 daktaro disertacijų santraukų (teorinio mokslinio postilio tekstai) iš visų mokslo sričių ir visų krypčių, išskyrus teologiją (pagal paieškos kriterijus teologijos krypties daktaro disertacijų santraukų nebuvo rasta). Žengus visus reikalingus tekstyno parengimo žingsnius, tyrimui buvo

¹⁰⁰ Lietuvos akademinės elektroninės bibliotekos portalas pasiekiamas adresu <https://elaba.lvb.lt>.

parengtą tikslinį mokslinio FS tekstyną sudaro 1025 tekstai (~ 20,2 mln. žodžių).

2.3.3 Publicistinio FS tekstynas

Šioje disertacijoje aprašomo tyrimo tikslams rengiant tikslinį publicistinio FS tekstyną, nuspręsta jo pagrindu imti ne DLKT, o projekto PASTOVU sukurtą delfi.lt tekstyną¹⁰¹. Projekto PASTOVU metu parengtą delfi.lt tekstyną sudaro naujesni, 2014–2016 m. naujienų portale Delfi¹⁰² publikuoti tekstai. Disertacijoje aprašomo tyrimo tikslams publicistinio FS tekstyno PST sudarymui nuspręsta atsitiktine tvarka, naudojant komandinės eilutės įrankį *shuf*¹⁰³, atrinkti po 1500 tekstų, ilgesnių nei 200 žodžių, iš 9 delfi.lt tekstyno teminių skilčių (temines kategorijas „projektai“ ir „DELFI FIT“ sudarė gerokai mažiau nei 1500 tekstų, tad šių kategorijų tekstų į tikslinį FS tekstyną nuspręsta neįtraukti).

Tad disertacijos tyrimui sudarytą pirminį tikslinį publicistinio FS tekstyną sudarė 13 500 tekstų. Galutinę šio tikslinio tekstyno versiją, atlikus paruošimo analizei žingsnius, sudarė 13 450 tekstų – 50 Sporto teminės kategorijos tekstų atkrito, nes po paruošimo analizei tapo per trumpi, t. y. juos sudarė mažiau nei 200 žodžių. Tikslinio publicistinio FS tekstų metaduomenys saugomi atskirai nuo pačio teksto, t. y. atskiruose failuose, išretintai parašytų žodžių nerasta, tad juos tereikėjo suskaidyti žodžiais ir suanotuoti morfologiškai, priskiriant morfologines pažymas. Žengus visus reikalingus tekstyno parengimo žingsnius, tyrimui buvo parengtą tikslinį publicistinio FS tekstyną sudaro 13 450 tekstų (~ 10,4 mln. žodžių).

¹⁰¹ Disertacijos autorė projekte dirbo jaunesniąja mokslo darbuotoja, todėl turėjo prieigą prie delfi.lt tekstyno; viešai yra prieinami šio tekstyno pagrindu parengti dažniniai n-gramų sąrašai (<http://resursai.mwe.lt/atsisiusti>).

¹⁰² Naujienų portalas pasiekiamas adresu <https://www.delfi.lt/>.

¹⁰³ Mokojoji medžiaga, kaip naudotis įrankiu, pasiekama adresu <https://www.ostechnix.com/the-shuf-command-tutorial-with-examples-for-beginners/>.

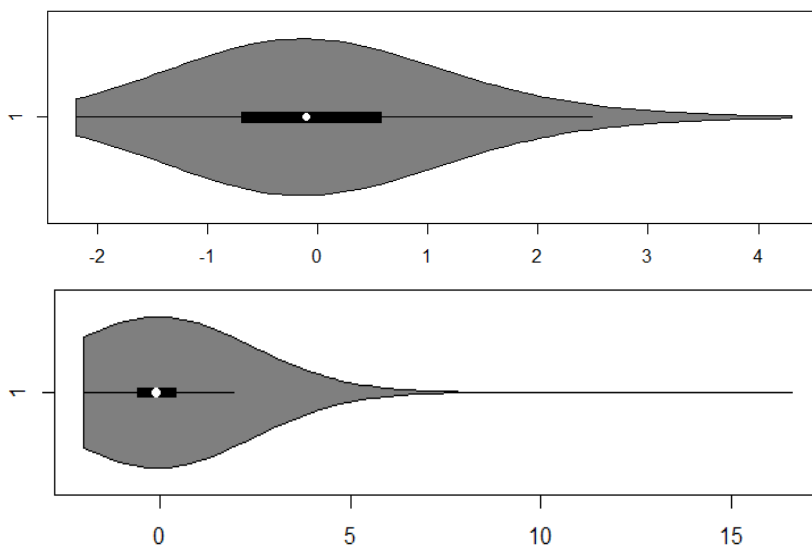
3. TYRIMO METODAI IR EIGA

3.1 Aprašomoji duomenų analizė: metodologinės pastabos

Analizuojant duomenis pirmiausia buvo atlikta aprašomoji duomenų analizė, kad būtų galima įvertinti tiriamų duomenų (kiekybinių indikatorių verčių kiekvienam tikslinių tekstynų tekstui) pagrindines statistines savybes (Casas, 2016). Aprašomoji duomenų analizė apima padėties, sklaidos ir kitų duomenų charakteristikų nustatymą.

Padėties charakteristikos apibūdina duomenų verčių didumą. Šiame tyrime analizuojamos 2 duomenų padėties charakteristikos: aritmetinis vidurkis ir mediana. Mediana kaip charakteristika įtraukta disertacijoje atliekamo tyrimo aprašomąją analizę, nes mažiau jautri išskirtims (neįprastai didelėms arba neįprastai mažoms vertėms) nei vidurkis.

Sklaidos charakteristikos apibūdina duomenų, t. y. indikatorių verčių, suskaičiuotų kiekvienam visų 3 tiriamų funkcinių stilių tekstynų tekstui, verčių išsisklaidymą, kuris žymi indikatorių raiškos įvairovę. Tyrime analizuojamos 3 sklaidos charakteristikos: standartinis nuokrypis, kvartilinis plotis ir variacijos (kitimo) koeficientas (Čekanavičius ir Murauskas, 2001, 41—42). Standartinis nuokrypis nurodo, kaip duomenys yra išsisklaidę apie vidurkį. Jei standartinio nuokrypio vertė yra artima 0, tai reiškia, kad duomenys yra mažai variantiški. Kitaip tariant, jei tam tikro kiekybinio indikatoriaus standartinis nuokrypio vertė yra artima 0, tai rodo mažą šio indikatoriaus raiškos įvairovę. Kita vertus, kuo didesnė standartinio nuokrypio vertė, tuo duomenys yra labiau išsisklaidę, t. y. raiškos įvairovė yra didesnė. Kaip galima matyti 4-ame paveiksle pateiktose smuiko (angl. *violin*) diagramose, viršuje pavaizduotiems duomenims būdingesnė didesnė duomenų sklaida nei pavaizduotiems apačioje, t. y. pirmojoje diagramoje „smuikas yra labiau išstėtas“.



4 pav. Smuiko diagramos duomenų savybių iliustravimui

Kvartilinis plotis (angl. *inter-quartile range*; toliau – KP) yra duomenų sklaidos charakteristika, pagrįsta skirtumu tarp 3-jo ir 1-jo kvartilų, t. y. atkarpų, reprezentuojančių 75 proc. ir 25 proc. kintamojo (šioje disertacijoje atliekamame tyrime kintamieji – funkcinių stilių kalbinę raišką žymintys indikatoriai) verčių sklaidos (Upton ir Cook, 1996, 55). 4-ame paveiksle KP vertės didumą žymi „smuiko“ viduje esančios juodos pastorinto linijos ilgis: viršuje pavaizduotiems duomenims priklauso didesnė KP vertė ir kartu didesnė duomenų sklaida, rodanti didesnę raiškos įvairovę, o apačioje – mažesnė KP vertė, rodanti mažesnę duomenų sklaidą bei mažesnę raiškos įvairovę. Ši charakteristika yra atspari išskirčių (itin didelių arba itin mažų verčių, išsiskiriančių bendrame duomenų kontekste (Čekanavičius ir Murauskas, 2001, 36)) sukeltam duomenų išsikreipimui, todėl laikoma patikima sklaidos charakteristika. Kuo kvartilinio pločio vertė didesnė, tuo duomenys labiau išsiskleidę vidurkio atžvilgiu, o disertacijoje atliekamo tyrime kontekste tai rodytų tam tikro indikatorius raiškos įvairovę atitinkamame funkciniame stiliuje. Kvartilinis plotis dažnai taikomas išskirčių radimui. Išskirtimis laikomos duomenų rinkinio reikšmės, kurios yra mažesnės nei $Q_1 - 1.5 KP$ arba didesnės nei $Q_3 + 1.5 KP$ (Lapinskas, 2003,

9)¹⁰⁴. Kvartilinis plotis išskirčių radimui taikytas ir šioje disertacijoje atliekamame tyrime.

Variacijos koeficientas apibūdina duomenų rinkinio reikšmių išsisklaidymo laipsnį. Tai yra santykinis dydis. Pavyzdžiui, variacijos koeficiento reikšmė 0,22 nurodo, kad standartinis nuokrypis yra 22 % aritmetinio vidurkio reikšmės. Jei koeficiento reikšmė būtų artima 0, tai rodytų, kad kintamojo reikšmės yra artimesnės vidurkiui, t. y. jei tam tikro kiekybinio indikatoriaus variacijos koeficiento vertė artima 0, tai rodo mažą raiškos įvairovę.

Be charakteristikų, priklausančių šioms dviem – padėties ir sklaidos – pagrindinėms grupėms, disertacijoje aprašomo tyrimo glaustam duomenų apibūdinimui dar taikyti asimetrijos ir eksceso koeficientai. Asimetrijos koeficientas (angl. *skewness*) yra statistinė duomenų charakteristika, apibūdinanti duomenų asimetriškumą. Kai asimetrijos koeficiento reikšmė lygi 0, duomenų dažnių kreivė yra simetriška. Jei ši reikšmė yra didesnė už 0, tai reiškia, kad dažnesnės duomenų reikšmės susitelkusios ties didesnėmis vertėmis, t. y. duomenys pasižymi tuo, kad jų didesnės vertės yra dažnesnės (Field, Miles ir Field, 2012, 19–21). Ši situacija pavaizduota 4-ojo paveikslo apatinėje diagramoje: „smuikas“ yra labiau „išstemtas“ dešinėn, o tai ir iliustruoja, kad tarp diagramoje pavaizduotų duomenų didesnės vertės, žyminčios didesnę tam tikro indikatoriaus raišką, yra dažnesnės nei mažesnės. Kai koeficiento reikšmė mažesnė nei 0, dažniausios duomenų reikšmės yra susitelkusios ties mažesnėmis vertėmis, o retesnės – ties didesnėmis, t. y. duomenyse dažniau pasitaiko mažesnių verčių nei didesnių (Field, Miles ir Field, 2012, 19–21).

Eksceso (angl. *kurtosis*) koeficientas yra statistinė duomenų rinkinio charakteristika, leidžianti įvertinti duomenis išskirčių (itin mažų arba itin didelių verčių, išsiskiriančių bendrame duomenų kontekste) atžvilgiu (Westfall, 2014). Jei duomenų pasiskirstymas atitinka normaliąją kreivę, t. y. jie yra simetriškai išsidėstę apie vidurkį, eksceso koeficiento vertė būtų 3 (Byrne, 2010, 103)). Didesnės arba mažesnės nei 3 šio koeficiento vertės rodo, kad išskirčių yra daugiau arba mažiau nei tai būdinga normaliajai kreivei.

¹⁰⁴ Q_1 – 1-asis kvartilis, kitaip 25% mažiausių duomenų verčių, 4-ame paveiksle jis vaizduojamas nuo diagramos pradžios kairėje iki per diagramos vidurį nubrėžto juodo pastorinto brūkšnelio pradžios; Q_3 – 3-ias ketvirtis duomenų verčių, 4-ame paveiksle jis vaizduojamas nuo „smuike“ nubrėžtame juodame pastorintame brūkšnelyje padėto balto taško, žyminčio medianą, iki šio brūkšnelio pabaigos.

Išskirčių įvertinimas yra svarbus tolesniam tyrimui – koreliacijos koeficiento (žr. 3.3 sk.) ir dispersinės analizės metodų pasirinkimui (žr. 3.5 sk.).

Aprašomoji analizė atlikta visiems 3-iems disertacijoje tiriamiems FS tekstinams, siekiant susisteminti šių tekstynų tekstams suskaičiuotų kalbinės raiškos požymius žyminčių 9 indikatorių (VŽFI, a , R_1 , RR_{mc} , MATTR, TK, A, D ir VA), naudotų tyrime, duomenis. Tai pirmasis tyrimo etapas, kurio rezultatai turėjo įtakos tolesnių tyrimo metodų pasirinkimu bei leido daryti pirmines išvadas apie FS požymius ir jų raišką.

3.2 Koreliacinė analizė: metodologinės pastabos

Duomenų analizės antrasis etapas – koreliacinė analizė. Koreliacinė analizės tikslas šioje disertacijoje atliekamame tyrime – įvertinti kalbinę raišką žyminčių kiekybinių indikatorių tarpusavio ryšius atskirų FS kontekste. Dažnai analizuojant kiekybinių kintamųjų reikšmių kitimą reikia atsakyti į klausimą apie stebimų kintamųjų priklausomumą ar nepriklausomumą bei ryšio tarp dviejų kintamųjų tendencijas. Pavyzdžiui, didėjant kintamojo X reikšmėms, kintamojo Y reikšmės didėja ar mažėja? Kintamųjų priklausomybės stiprumą vertina koreliacijos koeficientas (Field, Miles ir Field, 2012, 210–212). Tikrindami hipotezę apie duomenų koreliacijos koeficiento reikšmę, lygią nuliui, atsakome į klausimą apie kintamųjų priklausomybės statistinį reikšmingumą (Čekanavičius ir Murauskas, 2001, 165–167). Primintina, kad koreliacijos koeficientas nenusako koreliacijos priežasties (Field, Miles ir Field, 2012, 212). Du kintamieji X ir Y gali stipriai koreliuoti, t. y. turėti stiprų tarpusavio ryšį, dėl trijų priežasčių (Janilionis, Morkevičius ir Rauleckas, 2008): 1) kintamasis X daro poveikį kintamajam Y; 2) kintamasis Y daro poveikį kintamajam X; 3) abu kintamieji X ir Y yra veikiami trečio kintamojo. Tad koreliacinės analizės metu rastas ryšys tarp kintamųjų (šiuo atveju kalbinę raišką žyminčių kiekybinių indikatorių) interpretuojamas kaip asociacijos (ryšio), bet ne priežastingumo matas.

Atlikus koreliacinę analizę, gaunamos koreliacijos koeficiento reikšmės, leidžiančios įvertinti ryšį tarp 2 kintamųjų (čia – indikatorių). Šios reikšmės pasiskirsto intervale nuo -1 iki +1 (Field, Miles ir Field, 2012, 212). Tad koreliacijos koeficiento vertė +1 rodo, kad 2 kintamieji yra absoliučiai teigiamai koreliuoti, t. y. jei vieno kintamojo (indikatoriaus) reikšmės didėja, kito taip pat proporcingai didėja. Kita vertus, koreliacijos koeficiento vertė -1 reiškia absoliučią neigiamą koreliaciją tarp dviejų kintamųjų. Kitaip sakant, šiuo atveju vieno kintamojo (indikatoriaus) vertėms didėjant, kito –

proporcingai mažėja ir atvirkščiai. Jeigu koreliacijos koeficiento reikšmė lygi 0, vadinasi koreliacija tarp dviejų kintamųjų yra statistiškai nereikšminga (Field, Miles ir Field, 2012, 212), t. y. jei vieno kintamojo reikšmės kinta, kito – išlieka bemaž nepakitusios. Tai rodo, kad abu kintamieji yra statistiškai nepriklausomi. 12-oje lentelėje pateikta apibendrinta informacija, kokie koeficientai rodo kurias koreliacijų reikšmes nuo statistiškai nereikšmingos iki labai stiprios.

Tad koreliacinė analizė yra skirta įvertinti statistinę sąsają (ryšį) tarp dviejų kintamųjų (Čekanavičius ir Murauskas, 2001, 97—98, 124—126, 165—169). O koreliacijos koeficientas leidžia išmatuoti koreliacijos stiprumą. Norint patikrinti, ar ryšys tarp kintamųjų yra statistiškai reikšmingas, tikrinamos statistinės hipotezės (Čekanavičius ir Murauskas, 2001, 165—169):

- H_0 : dviejų požymių reikšmės nėra susijusios (vadinamoji nulinė hipotezė);
- H_1 : dviejų požymių reikšmės yra susijusios (alternatyvioji hipotezė).

„Požymių reikšmės yra susijusios“ reiškia, kad žinodami vieno požymio reikšmes gauname informacijos ir apie kito požymio reikšmes. Jei reikšmingumo lygmuo $\alpha=0,05$ (t. y. egzistuoja mažesnė nei 5 proc. tikimybė, kad nustatytas skirtumas tarp požymių buvo tik atsitiktinumas), tai, kai $p < 0,05$, laikoma, kad požymiai yra susiję (priimama H_1 , kad sąryšis statistiškai reikšmingas); kai $p \geq 0,05$, teigti, kad požymiai yra susiję, negalima (neatmetama H_0 , bet sakoma, kad ryšys tarp požymių yra statistiškai nereikšmingas) (Field, Miles ir Field, 2012, 212).

12 lentelė. Koreliacijos koeficiento reikšmės (pagal Čekanavičių ir Murauską (2001, 126))

Preliminarus koreliacijos vertinimas	Koreliacijos koeficiento reikšmės
koreliacija nereikšminga	0
labai silpna	[-0.2; 0] arba (0; 0.2]
silpna	[-0.4; -0.2] arba (0.2; 0.4]
vidutinė	[-0.7; -0.4] arba (0.4; 0.7]
stipri	[-0.9; -0.7] arba (0.7; 0.9]
labai stipri	[-1.0; -0.9] arba (0.9; 1.0]

Atliekant koreliacinę analizę kyla koreliacijos koeficiento pasirinkimo klausimas. Jei duomenys atitinka normaliąją kreivę, t. y. jie yra simetriški duomenų vidurkio atžvilgiu, nėra aiškių išskirčių (t. y. itin didelių arba mažų reikšmių, išsiskiriančių iš bendro duomenų konteksto) (Field, Miles ir Field, 2012, 219), ir duomenų rinkinys yra pakankamai didelis (pvz., kintamojo

duomenų ar kintamųjų grupės duomenų dydis N yra didesnis nei 25 stebėjimai (Kitchen, 2009)), tuomet tiriant duomenis pasirenkamas Pirsono (Pearson) koreliacijos koeficientas. Jei duomenų yra pakankami daug, bet jiems būdingos išskirtys, tuomet taikomas Spirmeno (Spearman) koreliacijos koeficientas (Mukaka, 2012). Mažoms duomenų imtims paprastai naudojamas Kendalo (Kendall) τ (tau) koreliacijos koeficientas (Field, Miles ir Field, 2012, 225). Šioje disertacijoje tiriami AST, MST ir PST tekstynai yra pakankamai dideli, tačiau, kadangi lingvistiniai duomenys dažnai yra nutolę nuo normaliosios kreivės (Levshina, 2015, 18), tiriant ryšius tarp indikatorių naudojamas Spirmeno koreliacijos koeficientas.

Kaip minėta, Spirmeno koreliacijos koeficientas, taikomas, kai duomenų pasiskirstymas neatitinka normaliosios kreivės (duomenys nėra simetriški vidurkio atžvilgiu, nėra „varpo“ formos) ir (ar) kai duomenyse gausu išskirčių, yra ranginis koreliacijos koeficientas (Field, Miles ir Field, 2012, 223). Kitaip sakant, ryšio tarp dviejų kintamųjų įvertinimui naudojamos ne kintamųjų reikšmės, o jų rangai. Tai yra, tiriamų 2 kintamųjų reikšmės išdėstomos pagal dydį, ir Spirmeno koreliacijos koeficientas įvertina ryšį tarp šių didėjimo ar mažėjimo tvarka išrikiuotų kintamųjų reikšmių sekų. Tad skaičiuojant šį koeficientą lyginami sekų rangai.

Iš esmės Spirmeno koreliacijos koeficientas yra tas pats Pirsono koreliacijos koeficientas, tik apskaičiuotas ne kintamųjų reikšmėms, o jų rangams (Čekanavičius ir Murauskas, 2002, 33). Taip pat, kaip ir pastarajam, Spirmeno koreliacijos koeficiento reikšmės pasiskirsto intervale nuo -1 iki +1, kur -1 žymi labai stiprų neigiamą ryšį, o +1 – labai stiprų teigiamą ryšį. Apibendrinta informacija, kokie koeficientai rodo kurias koreliacijų reikšmes nuo statistiškai nereikšmingos iki labai stiprios, pateikta 12-oje lentelėje.

Šioje disertacijoje, atliekant FS kalbinės raiškos požymių kiekybinį tyrimą, koreliacinė analizė taikyta ryšių tarp indikatorių, žyminčių FS kalbinę raišką, radimą kiekvienam FS. Tad koreliacinė analizė atlikta visiems tiriamiems FS tekstynams atskirai. Šis metodas leido įvertinti kiekvieno iš 9 indikatorių, naudotų tyrime (VŽFI, a , R_1 , RR_{mc} , MATTR, TK, A, D ir VA) tarpusavio ryšius tiriamų FS kontekste.

3.3 Pagrindinių komponenčių analizė: metodologinės pastabos

Duomenų analizės trečiasis etapas – pagrindinių komponenčių analizė, atlikta siekiant pateikti atsakymą, kurie indikatoriai yra svarbiausi apibūdinant kiekvieno tiriamo funkcinio stiliaus kalbinę raišką atskirai. Pagrindinių

komponenčių analizė (angl. principal component analysis, toliau – PKA) yra klasikinis statistikos metodas. Pagrindinė PKA idėja – sumažinti duomenų matmenų skaičių atliekant tiesinę transformaciją kartu išsaugant didžiąją dalį duomenų variantiškumo (Smith, 2002). PKA – tiesinė duomenų transformacija, t. y. duomenų (indikatorių verčių) išsidėstymo ašys pasukamos taip, kad viena šių ašių paaiškintų didžiausią duomenų variacijos dalį. Šioji ašis vadinama pirmąja komponente. Norint išskirti bendruosius duomenis paaiškinančius faktorius, pasilieka tik dalis pagrindinių komponenčių (Jolliffe, 2002). Visos pagrindinės komponentės apibūdina visą bendrą kintamųjų dispersiją. Pagrindinių komponenčių yra tiek, kiek yra kintamųjų. Pirmosios komponentės nusako didžiąją dalį visos duomenų dispersijos. Tai reiškia, kad jos gana tiksliai apibūdina kintamųjų (indikatorių) priklausomybės struktūrą (Janilionis, Morkevičius ir Rauleckas, 2008, 171), tad gali būti panaudotos bendrų faktorių skaičiavimui, t. y. geriausiai atskirų FS kalbinę raišką apibūdinančių indikatorių išskyrimui.

PKA disertacijoje atliekamo tyrimo atveju atrodytų taip: įsivaizduokime vieno kurio FS indikatorių vertes kiekvienam tekstui. Kadangi yra 9 indikatoriai, vadinasi, turėsime devynmačius duomenis (kiekvieno indikatoriaus duomenų vertės išdėstomos atskiroje ašyje). Duomenis standartizuojame, kad galėtume juos lyginti vienus su kitais. Tada, nekreipdami dėmesio į tai, kurios vertės kuriems indikatoriums priklauso, ieškome išdėstytuose duomenyse ašies (krypties), kurioje būtų didžiausia dispersija, ir atitinkamai pasukame duomenis ta kryptimi. Ši kryptis vadinsis pirmąja pagrindine komponente. Kadangi joje esantiems duomenims būdinga didžiausia dispersija, lyginant su likusiais, vadinasi, pirmajai komponentei priskirtos indikatorių vertės apibūdina didžiausią dalį duomenų sklaidos, t. y. didžiausią dalį šio vieno kurio FS kalbinės raiškos. Taip yra todėl, kad pirminiuose duomenyse vienoje ašyje buvo išdėstyti tik vieno indikatoriaus duomenis, o pasuktų (transformuotų) duomenų vienoje ašyje, t. y. pagrindinėje komponentėje, pateikiama indikatorių verčių kombinacija t. y. daugiau nei vieno indikatoriaus vertės.

Antroji ir tolesnės pagrindinės komponentės randamos panašiu principu: likusiuose duomenyse (atėmus tai, kas jau priklauso pirmajai pagrindinei komponentei, o jei ieškoma tolesnių komponenčių – analizuojami duomenys, neaprepti prieš tai ėjusių komponenčių) vėl ieškoma krypties, kurioje duomenų verčių dispersija yra didžiausia ir atitinkamai transformuojame arba pasukame duomenų išsidėstymo ašis. Kadangi komponentės talpina indikatorių verčių kombinacijas, tai leidžia geriau apibūdinti FS kalbinę raišką nei šia raišką tiriant pavieniais kiekybiniais požymiais. Tad PKA leidžia

išskirti kintamuosius (indikatorius), apibūdinančius didžiausią dalį FS kalbinės raiškos ir kartu atmesti tuos, kurių indėlis nėra reikšmingas¹⁰⁵.

Disertacijoje atliekant FS kalbinės raiškos požymių kiekybinę analizę, PKA metodas taikytas svarbiausių indikatorių, žyminčių FS kalbinę raišką, radimą kiekvienam FS. Tad PKA atlikta visiems tiriamiems funkcinių stilių tekstynams atskirai. Šis metodas leido įvertinti kiekvieno iš 9 indikatorių, naudotų tyrime (VŽFI, a , R_1 , RR_{mc} , MATTR, TK, A, D ir VA), indėlį ir svarbą apibūdinant tiriamus FS.

3.4 Dispersinė analizė: metodologinės pastabos

Taikant PKA metodą, tirti geriausiai atskirus FS reprezentuojantys indikatoriai bei jų kombinacijos. Kadangi FS yra bendros funkcinės stilistikos sistemos dalis, kitame tyrimo etape tiriami AST, MST ir PST tekstų tarpusavio panašumai ir skirtumai. Tam taikoma dispersinė analizė, leidžianti priklausomojo kintamojo skirtingas reikšmes paaiškinti vieno ar kelių nepriklausomų kintamųjų veikimu (Janilionis, Morkevičius ir Rauleckas 2008, 99), t. y. pagal indikatorių raišką įvertinti tiriamų FS panašumus ir skirtumus. Atsižvelgiant į tekstynų specifiką (išskirčių gausą, skirtingą skaičių tekstų, sudarančių atskirus tekstynus), šiame žingsnyje taikomi neparametriniai statistikos metodai (Field, Miles ir Field, 2012, 654). Taikant neparametrinius metodus, duomenims (čia – indikatorių vertėms) netaikomos normalumo, variacijos homogeniškumo ir kitos prielaidos, ribojančios tradicinių parametrinių testų, pavyzdžiui, daugiafaktorinės dispersinės analizės (MANOVA), kuri, jei duomenys atitiktų reikiamas prielaidas, būtų buvusi pasirinkta FS tarpusavio panašumų ir skirtumų ištyrimui, taikymą (Ellis, Burchett, Harrar ir Bathke, 2017). Neparametriniai testai paprastai yra

¹⁰⁵ Matematiškai trumpai PKA procedūrą galima nusakyti taip: 1) išskiriama pirmoji komponentė, paaiškinanti didžiausią dalį pradinių kintamųjų dispersijos, ir apskaičiuojama jos svorių vektorių atitinkanti tikrinė reikšmė (angl. *eigen value*). Ji parodo, kokią dalį dispersijos ši komponentė paaiškina. Tuomet ieškoma antrosios pagrindinės komponentės, kuri paaiškina didžiausią dalį likusios dispersijos ir nekoreliuoja su pirmąja, ir randama jos tikrinė reikšmė (parodo, kokią dalį pradinių duomenų variacijos paaiškina kiekviena komponentė (Janilionis, Morkevičius ir Rauleckas, 2008: 171)). Ši procedūra tęsiama tol, kol išskiriamos visos komponentės ir visa kintamųjų dispersija tampa jomis paaiškinama, t. y. sudėjus visas pagrindinių komponentių tikrines reikšmes, jos atitinka bendrą tiriamų duomenų kintamųjų dispersiją.

ranginiai, t. y. mažiausiai testo rezultato vertei duodamas mažiausias rangas (vieta pagal vertės dydį), pavyzdžiui, 1, sekančiai pagal mažumą (didesnei, nei mažiausia vertė) – 2 ir t. t. (Field, Miles ir Field, 2012, 654). Taip didžiausios testo vertes atitinka didžiausi rangai, o mažiausias – mažiausi.

Atliekant panašumų ir skirtumų tarp tiriamų funkcinių stilių analizę, visų pirma tirta, ar šie FS pagal 9 pasirinktus indikatorius tarpusavyje apskritai skiriasi. Tam pasirinktas Kruskal'o-Wallis'o testas (Field, Miles ir Field, 2012, 674—677; Mangiafico, 2016, 248—249). Kruskal'o-Wallis'o testas iš esmės yra parametrinės vienfaktorinės dispersinės analizės (ANOVA) atitinkmuo (Field, Miles ir Field, 2012, 674), naudojamas, kai lyginamos daugiau nei 2 nepriklausomos imtys (Mangiafico, 2016, 248). Disertacijos tyrimo kontekste šias imtis atitinka 3 tekstynai: AST, MST ir PST. Panašumų ir skirtumų laipsnio tarp tiriamų FS apskritai įvertinimui naudotos 2 statistikos: *epsilon* kvadratas (Tomczak ir Tomczak, 2014) bei Vargha ir Delaney A (VDA) rodiklis (Vargha ir Delaney, 2000). Pirmoji statistika – *epsilon* kvadratas – buvo skirtas įvertinti, kaip tiriamieji FS apskritai skiriasi pagal atskirus indikatorius, o VDA leido išskirti pagal atskirus indikatorius labiausiai tarpusavyje besiskiriančias FS poras.

Kruskal'o-Wallis'o testas leido įvertinti tik bendrus panašumus ir skirtumus tarp tiriamų FS, t. y. leido tik padaryti išvadą apie skirtumų buvimą arba nebuvimą. Tad tolesnėje tyrimo eigoje taikytas Dunn'o testas (Dinno, 2015), kuriuo įvertinti skirtumai tarp FS porų atskirų indikatorių atžvilgiu. Kadangi tyrime buvo atliekami daugybiniai palyginimai (FS poros lyginamos pagal kiekvieną tyrime naudojamą indikatorių), *p*-vertės buvo koreguojamos Bonfferoni metodu (Abdi, 2007). Indikatorių verčių tiriamiems FS tikimybinės tendencijos, t. y. kurie tiriami FS išsiskiria didesnėmis ar mažesnėmis vertėmis, rodančiomis didesnę ar mažesnę konkretaus FS požymio raišką, tirtos pasitelkus *R* statistinio paketo bibliotekos **nlmv** funkciją **nonpartest()** (Ellis, Burchett, Harrar ir Bathke, 2017). Ši funkcija leidžia atlikti neparametrinius testus išvadoms apie daugiamatį duomenų tarpusavio skirtumus gauti. Paleidus **nonpartest()** funkciją, pirmiausia skaičiuojamos pasirinktos neparametrinės statistikos (galimi 4 neparametriniai parametrinių testų variantai: ANOVA tipo neparametrinis testas, Lawley-Hotelling tipo neparametrinis testas, Bartlett-Nanda-Pillai tipo neparametrinis testas ir Wilks'o Lambda tipo neparametrinis testas), tada skaičiuojami neparametriniai santykiniai poveikio įverčiai (angl. *relative treatment effects*) kiekvienam kintamajam.

Šie įverčiai apibrėžiami kaip tikimybė, kad atsitiktinai parinktas stebėjimas, priklausantis grupei *k*, turės didesnę poveikį nei atsitiktinai parinktas stebėjimas iš viso tyrimui naudojamų duomenų rinkinio (Ellis,

Burchett, Harrar ir Bathke, 2017). Kitaip sakant, jei indikatoriaus VŽFI santykinis poveikio įvertis AST lygus 0,85076, tai rodo, kad tikimybė, jog atsitiktinai parinktam AST tekstui bus būdinga didesnė VŽFI vertė nei atsitiktinai parinktam tekstui iš visų 3 tiriamųjų tekstynų kartu, yra 0,85076 (arba 85,076 %). Tad šis pavyzdys rodo, jog administraciniam stiliui, lyginant su kitais tiriamais FS, būdingesni suprantamumo požiūriu sudėtingesni tekstai. Tad neparimetriniai santykiniai poveikio įverčiai leido nustatyti atskirų indikatorių verčių tendencijas FS požymių raiškai įvertinti ir, remiantis šiomis tendencijomis, ištirti tiriamų FS tarpusavio panašumus ir skirtumus.

3.5 Tyrimo eiga

Tiriamųjų FS (administracinio, mokslinio ir publicistinio) kalbinės raiškos tyrimą sudarė šie etapai:

1. Pirmame tyrime etape kiekvienam kiekvieno tikslinio FS tekstyno tekstui apskaičiuojamos pasirinktų 9 kiekybinių indikatorių vertės.
2. Tada atliekama aprašomoji (pirminė) duomenų analizė, skirta įvertinti indikatorių verčių bazinės statistines savybes kiekvienam tiriamajam tekstui.
3. Kitame etape įvertinamos netipiškos indikatorių vertės ir jų mastas bei atliekama koreliacinė analizė ryšių tarp indikatorių nustatymui kiekvieno tikslinio tekstyno atžvilgiu.
4. Paskui naudojama pagrindinių komponentų analizė (PKA) svarbiausių indikatorių, apibūdinant kiekvieno tiriamo FS kalbinę raišką atskirai, išskyrimui.
5. Galiausiai atliekama dispersinė analizė tiriamųjų FS tarpusavio skirtumų ir panašumų nustatymui.

4. ATSKIRŲ FS KALBINĖS RAIŠKOS POŽYMIŲ ANALIZĖ

4.1 Administracinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė

4.1.1 Aprašomoji (pirminė) administracinių tekstų (AST) požymių analizė

Siekiant susipažinti su tiriamaisiais duomenimis – administraciniais tekstais – ir jų savybėmis, buvo atlikta aprašomoji arba pirminė duomenų analizė. Tirtos statistinės charakteristikos leidžia daryti pirmines išvadas apie tiriamo objekto, savybes (Čekanavičius ir Murauskas, 2001, 25). Tolesniuose šio skyriaus poskyriuose pristatomos kiekvieno indikatoriaus, naudoto AST kalbinės raiškos požymių analizei, pirminės charakteristikos. Šios charakteristikos apskaičiuotos įvertinant indikatorių vertes kiekvienam AST tekstui. Bazinių statistinių charakteristikų santrauka kiekvienam tiriamam indikatoriui pateikiama 13-oje lentelėje.

13 lentelė. AST: bazinių statistinių charakteristikų santrauka¹⁰⁶

Indikatorius	Vidurkis	Mediana	Standartinis nuokrypis	Variacijos koef.	Kvartilinis plotis	Asimetrinės koef.	Eksceso koef.
VŽFI	7,426	7,397	0,376	0,051	0,487	0,41	4,4
<i>a</i>	7,900	7,728	2,462	0,312	3,210	1,10	9,8
R ₁	0,8097	0,8257	0,080	0,098	0,095	-1,38	6,2
RR _{mc}	0,9465	0,9503	0,022	0,023	0,025	-1,92	10,6
MATTR	39,77	40,58	3,679	0,093	4,084	-1,59	6,9
TK	0,6313 1	0,57505	0,303	0,480	0,407	0,74	3,4
A	0,6743 9	0,68932	0,149	0,221	0,189	-0,62	3,8

¹⁰⁶ Lentelėje pateiktų duomenų statistiniam pasiskirstymui grafiškai iliustruoti galima naudoti QQ (angl. *quantile-quantile*) diagramas (Field, Miles ir Field, 2012, 171-172), vis dėlto disertacijos autorei duomenų pateikimas lentele pasirodė parankesnis.

D	0,3256	0,3107	0,149	0,457	0,189	0,62	3,8
VA	12,468	6,261	34,713	2,784	2,294	8,42	87,9

Vidurkis ir mediana žymi „tipiškas“ kiekvieno indikatoriaus vertes, standartinis nuokrypis, variacijos koeficientas ir kvartilinis plotis leidžia įvertinti indikatorių verčių sklaidą. Asimetrijos koeficientas leidžia įvertinti, kokios konkrečiau indikatoriaus vertės yra dažnesnės – didesnės ar mažesnės. Eksceso koeficientas yra skirtas išskirčių duomenyse įvertinimui bei tam tikrų netipiškų atvejų pagal tam tikro indikatoriaus raišką masto nustatymui, o tai turi įtakos tolesnių tyrimo metodų pasirinkimui. Tolesniuose poskyriuose pateikiama tyrime naudotų indikatorių detalūs aprašomosios analizės rezultatai AST tekstams (4.1.1.1-4.1.1.8 sk.)

4.1.1.1 Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus savybės AST

VŽFI indikatorius dažnai naudojamas tekstų suprantamumui / sudėtingumui (angl. *readability*) įvertinti. AST tekstams VŽFI reikšmės pasiskirsčiusios intervale nuo 5,373 (mažiausia reikšmė, priklausanti tekstui *adm213-177.txt*, kurį sudaro Lietuvos Respublikos Ryšių ir informatikos ministerijos įsakymas „Dėl pašto paslaugų tarifų pakeitimo“) iki 9,209 (didžiausia reikšmė, priklausanti tekstui *adm131-25.txt*, kurio turinys – Lietuvos Respublikos Aukščiausiosios Tarybos Prezidiumo nutarimas „Dėl Lietuvos Respublikos Aukščiausiosios Tarybos aparato darbuotojų tarnybinių atlyginimų“). Kitaip tariant, ši tekstyną sudarančių individualių tekstų vidutinis žodžio ar žodžio formos ilgis yra nuo 5,373 iki 9,209 simbolių (plg.: MST – nuo 5,208 iki 11,536 simbolių (plačiau žr. 4.2.1 sk.), o PST – nuo 5,013 iki 7,945 simbolių (plačiau žr. 4.3.1 sk.)). 13-oje lentelėje pateikiamos bazinės VŽFI verčių AST charakteristikos. Kaip rodo 13 lentelė, indikatoriaus kiekvienam AST tekstui verčių vidurkio vertė (7,426) yra truputį didesnė nei medianos (7,397). Tai rodo, kad išskirčių, t. y. itin didelių arba mažų verčių, išsiskiriančių iš visų indikatoriaus verčių, jas išrikiavus pagal dydį, galima tikėtis didelių verčių pusę. Kitaip tariant, vidutiniškai itin ilgi žodžiai ir žodžių formos AST atžvilgiu, tikėtina, bus veikiau išimtis nei taisyklė.

Mažas, lyginant su vidurkiu, VŽFI verčių standartinis nuokrypis (0,376) rodo, kad šio indikatoriaus vertės yra artimos vidurkiui (Field, Miles ir Field, 2012, 40), t. y. teksto sudėtingumo suprantamumo prasme raiška AST tekstams yra gana panaši, tad šie tekstai tarpusavyje nelabai skiriasi. Variacijos koeficiento vertė (0,051) reiškia, kad standartinis nuokrypis sudaro

tik 5,1 proc. aritmetinio vidurkio reikšmės, tad tai vėl parodo mažą indikatoriaus verčių sklaidą (Dormann, 2017, 13). Tad VŽFI požiūriu AST tekstai tarpusavyje nedaug skiriasi. Pagal kvartilinio pločio vertę galima sakyti, kad AST tekstai vidutinio žodžio ar žodžio ilgio atžvilgiu tarpusavyje skiriasi mažiau nei MST ir PST tekstai (žr. 13, 20 ir 25 lenteles).

Asimetrijos koeficiento vertė VŽFI indikatoriumi yra teigiama (0,41), bet mažesnė nei 0,5. Tai rodo, kad indikatoriaus vertės yra pasiskirsčiusios beveik simetriškai (Bulmer, 1979, 63). Toks VŽFI reikšmių pasiskirstymas reiškia, kad vidutiniškai ilgesni nei vidurkis žodžiai ir žodžių formos yra šiek tiek dažnesni. Eksceso koeficiento vertė (4,4) rodo, kad VŽFI duomenyse turėtų būti daugiau išskirčių (išskirtys yra itin didelės arba itin mažos reikšmės, išsiskiriančios iš bendro duomenų konteksto (Čekanavičius ir Murauskas, 2001, 36)) nei būdinga normaliajai kreivei (Westfall, 2014). Kitaip sakant, tarp VŽFI verčių turėtų būti šiek tiek netipiškų verčių. Tai svarbu įvertinti renkantis tolesnius tyrimo metodus.

4.1.1.2 Indikatoriaus *a* savybės AST

Indikatorius *a* žymi labai dažnų (paprastai – funkcinių, nors nebūtinai) žodžių ir žodžių formų srities dydį tekstą sudarančių žodžių dažniniame sąrašė. Indikatoriaus bazinių statistinių savybių nustatymui atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 13-ioje lentelėje. Šio indikatoriaus reikšmės AST pasiskirstė intervale nuo 1,576 (priklauso tekstui *adm113-28.txt* – Lietuvos Respublikos Vyriausybės nutarimas „Dėl turto perdavimo“) iki 39,50 (priklauso tekstui *adm105-328.txt* – Lietuvos Respublikos Vyriausybės nutarimas „Dėl Lietuvos Respublikos administracinių teritorinių vienetų priskirtų pasienio ruožui sąrašo patvirtinimo“) (płg.: MST – 5,450 iki 25,429 (plačiau žr. 4.2.1 sk.), o PST – nuo 3,59 iki 43,22 (plačiau žr. 4.3.1 sk.). Kuo didesnė indikatoriaus *a* vertė, tuo mažesnė teksto dažniniame sąrašė labai dažnų žodžių sritis, o kuo *a* vertė mažesnė – tuo ši sritis yra didesnė. Kitaip sakant, didelės *a* vertės rodo, jog tekstui nebūdinga labai dažnų (paprastai – funkcinių) žodžių ir jų formų gausa, o mažos *a* vertės reiškia, jog, priešingai, tekstui labai dažnų žodžių ir žodžių formų gausa būdinga. Kaip rodo analizės duomenys, šio indikatoriaus verčių kiekvienam AST tekstui vidurkis (7,900) yra didesnis nei mediana (7,728). Tai rodo, kad indikatoriaus vertės išrikiavus pagal dydį, išskirčių galima tikėtis toje pusėje, kurioje rikiuojasi didelės vertės. Kitaip sakant, maža labai dažnų žodžių ir jų formų sritis AST yra veikiau išimtis negu dėsningumas.

Kaip matyti iš 13 lentelės, indikatoriaus reikšmių standartinio nuokrypis (2,462) yra didelis, lyginant su vidurkiu, o tai reiškia, kad *a* reikšmės yra labai išsisklaidžiusios (Field, Miles ir Field, 2012, 40). Tai rodo, kad AST tekstams būdinga šio indikatoriaus raiškos įvairovė, t. y. AST tekstai labai dažnų žodžių srities dydžio atžvilgiu reikšmingai skiriasi tarpusavyje. Variacijos koeficiento reikšmė (0,312) parodo, kad standartinis nuokrypis sudaro net 31,2 proc. aritmetinio vidurkio reikšmės, o tai patvirtina didelę duomenų sklaidą ir tekstų tarpusavio skirtingumą indikatoriaus *a* atžvilgiu. Pagal kvartilinį plotį galima spręsti, kad *a* verčių sklaida AST yra mažesnė nei MST, bet didesnė nei PST (žr. 13, 20, ir 25 lenteles). Kitaip sakant, AST tekstai pagal labai dažnų žodžių srities dydį tarpusavyje skiriasi mažiau nei MST, bet labiau nei PST.

14 lentelė. AST teksto *adm0-100.txt* dažninės lentelės fragmentas (pilka spalva žymi h-tašką)

Eil nr.	Žodis	Dažnis
1	ir	10
2	respublika	6
3	finansų	5
4	ploto	4
5	dienpinigių	4
6	gyvenamojo	4
7	į	4
8	nuomos	4
9	salos	4
10	gvinėja ¹⁰⁷	4
11	ministerijos	3
12	vykstantiems	3
13	komandiruotes	3
14	užsienio	3
15	naujoji	3

¹⁰⁷ Tekstyno parengimo analizei etape visos žodžius sudarančios raidės buvo paverstos mažosiomis, nes R paketas, kuris buvo naudojamas tyrime, didžiąja ir mažąja raidėmis parašytus žodžius laiko skirtingais (pavyzdžiui, „bėgo“ ir „Bėgo“ (sakinio pradžioje) būtų laikomi skirtingais žodžiais), o tai iškreiptų skaičiavimo rezultatus.

15 lentelė. AST teksto *adm81-74.txt* dažninės lentelės fragmentas (pilka spalva žymi h-tašką)

Eil. nr.	Žodis	Dažnis
1	ir	49
2	įmonių	30
3	įmonės	26
4	lietuvos	22
5	respublikos	22
6	akcinės	20
7	bendrovės	17
8	registravimo	15
9	turi	14
10	vietos	14
11	jų	12
12	rejestro	12
13	jeigu	12
14	ar	12
15	vardai	11

13 lentelėje galima matyti, kad asimetrijos koeficiento vertė a indikatoriumi yra teigiama ir didesnė nei 1 (1,10). Tai reiškia, kad administracinio stiliaus tekstams būdingesnės didesnės a vertės, t. y. šios didesnės vertės yra dažnesnės nei tai tipiškai normaliajai kreivei (Field, Miles ir Field, 2012, 176). Kitaip sakant, didesnės a vertės rodo, kad tekstui būdinga mažesnė labai dažnų žodžių sritis (žr. 14-ą lentelę), o mažesnės – kad tekstui tipiškesnė didesnė labai dažnų žodžių sritis (žr. 15-ą lentelę). Didesnė nei 3 eksceso koeficiento vertė (žr. 13 lentelę) rodo, kad indikatorius a duomenyse bus išskirčių (Westfall, 2014). Kitaip sakant, tai reiškia, kad tarp a turėtų būti netipiškų – itin didelių arba itin mažų, ir į tai reikės atkreipti dėmesį renkantis tolesnius tyrimo metodus.

4.1.1.3 R_1 indikatorius savybės AST

R_1 indikatorius paprastai naudojamas žodyno turtingumui matuoti (Kubát, Matlach ir Čech, 2014, 14). Dažniausiai žodyno turtingumas vertinamas skirtingų žodžių skaičiumi tekste (Hoover, 2003). Lietuvių kalbai būdinga

kaitybinių formų gausa, todėl ši žodyno turtingumo traktuotė nėra tinkama, nebent būtų tiriamos ne žodžių formos, o lemos, t. y. žodžių antraštinės formos (Rimkutė, 2006). Tad kadangi šioje disertacijoje FS kalbinės raiškos požymiams analizuoti pasitelktos kaitybinės žodžių formos, o ne lemos, šiuo atveju tiksliau sakyti, kad R_1 yra žodžių ir jų formų įvairovės indikatorius. Primintina, kad ši R_1 vertinama įvairovė yra retesnių (priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų įvairovė.

Siekiant pažinti indikatorius R_1 bazines statistines savybes, buvo atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 13-oje lentelėje. R_1 vertės administracinio stiliaus tekstams pasiskirstė nuo 0,2397 iki 0,9595 (plg.: MST – 0,5625 iki 0,9250 (plačiau žr. 4.2.1.3 sk.), o PST – nuo 0,5446 iki 0,9799 (plačiau žr. 4.3.1.3 sk.). O tai reiškia, kad buvo ir nedidelės žodžių ir jų formų įvairovės tekstų, ir didelės. Mažiausia žodžių ir jų formų įvairovė būdinga tekstui *adm35-27.txt* – Lietuvos Respublikos Energetikos ministerija įsakymas „Dėl licencijų išdavimo“ (R_1 vertė lygi 0,2397). Didžiausia žodžių ir jų formų įvairovė išsiskyrė *adm45-74.txt* – Lietuvos Respublikos Vyriausybės nutarimas „Dėl Lietuvos Respublikos alkoholio kontrolės įstatymo 44 straipsnyje numatytų ekonominių sankcijų taikymo atidėjimo“ (R_1 vertė lygi 0,9595). Šio indikatorius reikšmių vidurkis (0,8097) yra mažesnis nei mediana (0,8257) (13 lentelė). Tai rodo, kad išskirčių (netipiškai didelių arba mažų verčių) tikėtina rasti mažų verčių pusėj. Kitaip tariant, mažesnė žodžių ir jų formų įvairovė, tikėtina, bus išimtis, o ne tendencija.

Indikatorius verčių standartinis nuokrypis, lyginant su vidurkiu, yra vidutiniškai didelis (0,080), tad tai reiškia, kad R_1 reikšmės yra vidutiniškai artimos jo vidurkiui (Field, Miles ir Field, 2012, 40). Kitaip tariant, AST iš dalies turėtų linkti link didesnės žodžių ir žodžių formų įvairovės. Variacijos koeficiento vertė (0,098) yra didesnė nei 0,05, bet mažesnė nei 0,2, todėl galima manyti, jog indikatorius verčių variantiškumas ir kartu AST tekstų tarpusavyje skirtingumas yra vidutiniškas. Lyginant visus tiriamuosius tekstynus, nors R_1 verčių sklaida iš esmės yra vidutinė, tačiau vis tiek didesnė nei MST ir PST. Kitaip sakant, šio indikatorius atžvilgiu AST tekstai tarpusavyje skiriasi labiau nei MST ir PST (žr. 13, 20, ir 25 lenteles). Kvartilinio pločio vertė (0,095) taip pat rodo, kad pagal R_1 duomenis AST tekstai tarpusavyje skiriasi mažiau nei MST ir PST.

Asimetrijos koeficiento vertė R_1 indikatorius yra neigiama (-1,38), o tai rodo, kad AST tekstai, kuriems būdingesnė mažesnė (retesnių) žodžių ir jų formų įvairovė yra dažnesni nei tokie, kurių ši įvairovė didesnė. Be to, eksceso koeficiento vertė (6,2) parodo, kad R_1 duomenyse turėtų būti daugiau išskirčių, nei tai būdinga normaliajai kreivei (Westfall, 2014) ir to reikės nepamiršti atliekant tolesnį funkcinių stilių kalbinės raiškos tyrimą.

4.1.1.4 AST tekstų RR_{mc} indikatoriaus charakteristikos

RR_{mc} šios disertacijos tyrimo atveju žymi tekste vartojamų žodžių formų koncentraciją (Kubát, Matlach ir Čech, 2014, 16), t. y. kuo aukštesnė šio indikatoriaus reikšmė, tuo tekstui tipiškesnė didesnė žodžių ir jų formų koncentracija. RR_{mc} vertinamo tekste vartojamų žodžių koncentruotumo reikšmės pasiskirsto intervale nuo 0 iki 1. Indikatoriaus bazinių statistinių savybių santrauka, atlikus aprašomąją duomenų analizę, pateikta 13-oje lentelėje. AST individualių tekstų RR_{mc} reikšmės pasiskirstė nuo 0,7398 (mažiausia reikšmė) iki 0,9858 (didžiausia reikšmė) (plg.: MST – nuo 0,8884 iki 0,9779 (plačiau žr. 4.2.1.4 sk.), o PST – nuo 0,8321 iki 0,9932 (plačiau žr. 4.3.1.4 sk.)). Tekstas *adm213-177.txt* – Lietuvos Respublikos Ryšių ir informatikos ministerijos įsakymas „Dėl pašto paslaugų tarifų pakeitimo“ įgijo mažiausią RR_{mc} reikšmę. Tad šiam tekstui būdinga mažiausia žodžių ir jų formų koncentracija iš visų AST tekstų. Didžiausią RR_{mc} reikšmę įgavo tekstas *adm55-16.txt* – Lietuvos Respublikos Prezidento dekretas „Dėl apdovanojimo Lietuvos kariuomenės kūrėjų savanorių medaliu“. Taigi šiam tekstui būdinga, lyginant visus AST tekstus, didžiausia žodžių ir jų formų koncentracija. Indikatoriaus verčių kiekvienam AST tekstui vidurkis (0,9465) yra mažesnis nei mediana (0,9503). Tai rodo, mažų verčių pusėj (indikatoriaus vertes kiekvienam AST tekstui išrikiavus pagal didumą) galima tikėtis išskirčių – šiuo atveju, labai mažų RR_{mc} verčių, išsiskiriančių bendrame visų šio indikatoriaus verčių kontekste. Kitaip tariant, maža AST tekstų žodžių ir jų formų koncentracija turėtų būti greičiau išimtis negu dėsningumas.

Kaip rodo 13 lentelėje pateikti duomenys, indikatoriaus reikšmių standartinio nuokrypio vertė, lyginant su vidurkiu, yra maža (0,022), tad RR_{mc} reikšmės yra jam artimos, o tai reiškia mažą indikatoriaus įvairovę, t. y. AST tekstai žodžių ir žodžių formų koncentruotumo atžvilgiu tarpusavyje mažai skiriasi. Variacijos koeficiento šiam indikatoriumi vertė (0,023) taip pat rodo, kad AST tekstams būdingas panašus žodžių ir jų formų koncentruotumas. Kvartilinio pločio vertė RR_{mc} rodo, kad, nors AST tekstai tarpusavyje mažai skiriasi, tačiau vis tiek skirtingesni nei MST ir PST tekstai (žr. 13, 20, ir 25 lenteles).

Kaip matyti iš 13 lentelės, RR_{mc} verčių išsidėstymui būdinga didelė neigiama asimetrija. Kitaip sakant, AST tekstai, kuriems būdinga mažesnė žodžių ir jų formų koncentracija yra dažnesni. Eksceso koeficiento vertė (10,6) reiškia, kad duomenyse turėtų būti daugiau išskirčių nei tai būdinga normaliajai kreivei, t. y. labai didelių arba mažų verčių, kurios išsiskiria iš

bendro RR_{mc} verčių „rinkinio“. Į šių netipiškų verčių mastą atsižvelgiama pasirenkant metodus tolesniam funkcinį stilių kalbinės raiškos tyrimui.

4.1.1.5 MATTR indikatoriaus charakteristikos AST

MATTR – nuo teksto ilgio nepriklausomas informacijos srauto, įvertinamo pagal skirtingų žodžių ir žodžių formų skaičių tekste, matas. Duomenų analizė atskleidė, kad MATTR reikšmės AST pasiskirsčiusios intervale nuo 14,76 (mažiausia reikšmė) iki 48,17 (didžiausia reikšmė) (plg.: MST – nuo 33,44 iki 47,61 (plačiau žr. 4.2.1.5 sk.), o PST – nuo 26,94 iki 48,97 (plačiau žr. 4.3.1.5 sk.). Didžiausia MATTR reikšmė, žyminti didžiausią AST informacijos „srautą“ priklauso tekstui *adm14-12.txt* – Lietuvos Respublikos Seimo nutarimas „Dėl Seimo komitetų narių pavaduotojų patvirtinimo“. Šį tekstą sudaro 696 žodžių ir jų formos, iš jų net 492 – skirtingi. Mažiausią MATTR vertę, rodančią mažiausią informacijos srautą, įgavo tekstas *adm46-59.txt* – Lietuvos Respublikos įstatymas „Dėl apylinkių teismų teisėjų etatų skaičiaus“. Tekstą sudaro 368 žodžiai ir jų formos, iš jų tik 76 – skirtingi. Bazinės indikatoriaus statistines charakteristikas galima matyti 13-oje lentelėje. Indikatoriaus vidurkis AST tekstams yra mažesnis nei mediana. Tai rodo, kad išskirčių (itin didelių arba mažų verčių) galima tikėtis mažų verčių pusėj. Kitaip tariant, AST tekstams tipiškesnės yra didesnės MATTR vertės, žyminčios didesnę informacijos „srautą“.

Indikatoriaus verčių standartinis nuokrypis (3,679), lyginant su vidurkiu, yra vidutiniškai didelis, o tai rodo, kad MATTR reikšmės yra iš dalies išsisklaidžiusios. Kitaip tariant, toks rodiklis atskleidžia, kad AST tekstai informacijos srauto atžvilgiu tarpusavyje skiriasi vidutiniškai. Variacijos koeficiento reikšmė (0,093) patvirtina šią tendenciją. Kita vertus, nors MATTR verčių sklaida AST yra vidutiniška, tačiau AST tekstai tarpusavyje skiriasi labiau nei MST ir PST (žr. 13, 20, ir 25 lenteles). Kvartilinis plotis irgi nurodo šią tendenciją (žr. 13, 20, ir 25 lenteles).

Asimetrijos koeficiento vertė MATTR indikatoriui yra neigiama ir mažesnė už -1 (-2,090). Tai rodo didelę duomenų verčių asimetriją (Bulmer, 1979, 63). Tai reiškia, kad dažnesnės yra mažesnės MATTR vertės, kitaip – AST tekstams būdingesnis mažesnis informacijos srautas. Eksceso koeficiento vertė (6,9) rodo, kad indikatoriaus vertėms būdinga didesnė gausa išskirčių nei tai tipiška normaliajai kreivei (Westfall, 2014) ir į šių netipiškų verčių mastą reikės nepamiršti atsižvelgti kitame tyrimo etape.

4.1.1.6 AST teminės koncentracijos (TK) indikatoriaus charakteristikos

TK indikatorius vertina, kiek teksto raiška sukoncentruota į gvildenamą temą (Kubát, Matlach ir Čech, 2014, 48; Čech, 2016; Kubát ir Čech, 2016b) (plačiau žr. 1.4.3.1 sk.). TK reikšmės AST pasiskirsčiusios nuo 0,0505 (mažiausia reikšmė – mažiausia teminė koncentracija) iki 1,8833 (didžiausia reikšmė – didžiausia teminė koncentracija) (plg.: MST – nuo 0 iki 0,92857 (plačiau žr. 4.2.1.6 sk.), o PST – nuo 0 iki 72,4 (plačiau žr. 4.3.1.6 sk.)). Mažiausiai temiškai koncentruotas buvo tekstas *adm81-91.txt* – Lietuvos Respublikos Statybos ir urbanistikos ministerijos įsakymas „Dėl teritorinio planavimo ir inžinerinės įrangos projektavimo rekomendacijų apsaugai nuo smurto ir vandalizmo patvirtinimo“. Šiame tekste riba tarp labai dažnų ir retesnių žodžių ir žodžių formų yra ties 16,5 (t. y. ties šia pozicija yra h-taško vieta), virš šios ribos buvo 5 turinio žodžiai ir jų formos (*turi, gyventojų, namo, namų, rekomenduojama*). 3 iš šių 5 žodžių ir žodžių formų pozicijos dažninėje lentelėje buvo arti h-taško (*namo* – 13 pozicijoje, *namų* – 14 pozicijoje, *rekomenduojama* – 16). Turinio žodžiai ir jų formos, esančios arti minėtojo h-taško, lėmė, kad šis tekstas laikomas temiškai beveik neutraliu (jei TK vertė būtų 0, tai rodytų, kad tarp labai dažnų žodžių ir žodžių formų virš h-taško nėra įsiterpę retesnių (priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų, tad toks tekstas būtų laikomas temiškai neutraliu (Čech, Garabík ir Altmann, 2015)).

Temiskai koncentruočiausias buvo tekstas *adm17-64.txt* – Lietuvos Respublikos Aukščiausiosios Tarybos nutarimas dėl Lietuvos Respublikos laikinojo įstatymo „Dėl atsakomybės už prekybos taisyklių, prekių supirkimo, išvežimo ar išsiuntimo už Lietuvos Respublikos ribų tvarkos pažeidimus“ taikymo. Šiame tekste riba tarp labai dažnų ir retesnių žodžių ir jų formų dažninėje lentelėje buvo ties 5 pozicija (h-taško vieta), o 2 patys dažniausi žodžiai – turinio/teminiai (dažniausias – *Respublikos*, antras pagal dažnumą – *Lietuvos*). Šie 2 žodžiai, būdami pačioje dažninės lentelės viršuje, turėjo itin didelį teminį svorį, tad tai ir lėmė, kad tekstas laikomas temiškai itin koncentruotu.

Siekiant įvertinti TK indikatoriaus statistines savybes, buvo atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 13-oje lentelėje. Indikatoriaus verčių individualiems administracinio teksto tekstams vidurkis (0,63131) buvo didesnis nei mediana (0,57505). Tai rodo, kad itin didelių arba mažų verčių, išsiskiriančių iš indikatoriaus verčių visumos (kitaip – išskirčių), galima tikėtis didelių verčių pusėj. Kitaip tariant, itin didelė

teminė koncentracija administraciniuose tekstuose turėtų būti greičiau išimtis nei dėsningumas.

Kaip matyti iš 13 lentelės, lyginant su vidurkiu (0,63131), TK standartinis nuokrypis (0,303) yra didelis, o tai rodo, jog indikatorius vertės yra nuo jo nutolusios nuo vidurkio (Field, Miles ir Field, 2012, 40). Kitais žodžiais tariant, TK raiška AST įvairuoja, t. y. pagal šį požymį AST tekstai tarpusavyje labai skirtingi. Variacijos koeficiento reikšmė (0,480) patvirtina šią tendenciją. Kvartilinio pločio vertė (0,407) rodo, kad pagal teminę koncentraciją AST tekstai tarpusavyje skiriasi labiau nei MST ir PST (žr. 13, 20 ir 25 lenteles).

Asimetrijos koeficiento vertė rodo, kad TK indikatorius vertėms būdinga teigiama vidutinė asimetrija (koeficiento rodiklis 0,74 yra didesnis nei 0,5, bet mažesnis nei 1 (Bulmer, 1979, 63)). Kitaip sakant, AST tekstams būdingesnis didesnis teminis koncentruotumas nei vidurkis, t. y. didesnės nei vidurkis TK vertės yra dažnesnės. Eksceso koeficiento rodiklis, šiek tiek didesnis nei 3 (žr. 13 lentelę) rodo, kad duomenims būdingas truputį didesnis skaičius išskirčių (itin didelių arba itin mažų reikšmių) nei tai tipiška normaliajai kreivei (Westfall, 2014). Kitaip sakant, tarp TK verčių turėtų būti šiek tiek netipiškų – itin didelių arba itin mažų, ir į jų mastą atsižvelgiama pasirenkant metodus kitam tyrimo etapui.

4.1.1.7 AST tekstų indikatorių A ir D charakteristikos

A (aktyvumo) indikatorius yra paremtas idėja, kad teksto aktyvumą paprastai nusako veiksmazodžiai, o jam priešingą deskriptyvumą – būdvardžiai (Kubát, Matlach ir Čech, 2014, 43; Kubát, 2016, 77–78). Kaip rodo M. Kubát čekų kalbos tekstų tyrimai, romano, pasakos ir panašių tekstų A įvertis bus didesnis (daugiau veiksmazodžių lyginant su būdvardžiais) nei mokslinio tyrimo ar kelionių aprašyme (šiuose ir panašiuose tekstuose sutinkama daugiau būdvardžių nei veiksmazodžių) (Kubát, 2016, 80). Tai šio darbo autorei leido daryti prielaidą, kad administracinio tekstyno atveju šis indikatorius turės mažesnę įvertį. Šie indikatoriai yra vienas kitam priešingi, t. y. vieno jų raiškai didėjant, kito – mažėja ir atvirkščiai.

Tyrimas atskleidė, kad A reikšmės administracinio stiliaus tekstyno tekstuose pasiskirstė nuo 0,0342 (mažiausia reikšmė, žyminti mažiausiai dinamišką tekstą) iki 1 (didžiausia reikšmė, žymi dinamiškiausius tekstus) (płg.: MST – nuo 0,2785 iki 0,8376 (plačiau žr. 4.2.1.7 sk.), o PST – nuo 0,2727 iki 0,9619 (plačiau žr. 4.3.1.7 sk.)). Tuo tarpu D vertės AST tekstams yra nuo 0 (mažiausia reikšmė, žyminti mažiausią deskriptyvumo raišką) iki

0,9658 (didžiausia reikšmė, žyminti mažiausią deskriptyvumo raišką) (plg.: MST – nuo 0,1624 iki 0,7215 (plačiau žr. 4.2.1.7 sk.), o PST – nuo 0,0381 iki 0,7273 (plačiau žr. 4.3.1.7 sk.)). Mažiausia aktyvumo raiška (ir kartu – didžiausia deskriptyvumo) buvo būdinga tekstui *adm213-212.txt* – „Valstybinių studijų ir mokymo įstaigų bei valstybės pripažintų nevalstybinių švietimo įstaigų sąrašas“. Jame yra 113 būdvardžių ir tik 4 veiksmožodžiai. Didžiausia aktyvumo raiška (ir kartu – mažiausia deskriptyvumo) buvo tipiška 69 tekstams, kuriuose nebuvo nei vieno būdvardžio. Didžiąją dalį šių tekstų sudarė Lietuvos Respublikos Prezidento dekretai dėl pilietybės atsisakymo arba suteikimo (iš esmės – ilgi vardų ir pavardžių sąrašai). Likę tekstai taip pat buvo sąrašai, pavyzdžiui, Lietuvos Respublikos Ministro Pirmininko potvarkis „Dėl atsargos karininkų laipsnių“, Lietuvos Respublikos įstatymas „Dėl apylinkių teismų teisėjų etatų skaičiaus“, Lietuvos Respublikos Ryšių ir informatikos ministerijos pranešimas „Dėl licencijų statyti ir eksploatuoti radijo ieškos ir standartų tinklus išdavimo“ ir keletas kitų.

Indikatorių A ir D statistinių savybių įvertinimui buvo atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 13-oje lentelėje. Indikatoriaus A verčių visiems administracinio stiliaus tekstyno tekstams vidurkis yra šiek tiek mažesnis nei mediana, o indikatoriaus D – šiek tiek didesnis. Tai rodo, kad A ir D vertės pasiskirsčiusios apie vidurkį beveik simetriškai, o išskirčių, t. y. itin didelių arba mažų verčių, jas išrikiavus pagal dydį, aktyvumo atžvilgiu galima tikėtis mažų verčių pusėj, o deskriptyvumo – didelių. Kitaip tariant, itin maža aktyvumo arba itin didelė deskriptyvumo raiška administraciniuose tekstuose turėtų būti išimtiniai atvejai.

Kaip matyti iš 13 lentelės, nemažas, lyginant su vidurkiu, A ir D verčių standartinis nuokrypis rodo, kad šio indikatorių vertės yra nutolusios nuo vidurkio (Field, Miles ir Field, 2012, 40). Tai rodo, kad A ir D vertės yra išsisklaidžiusios, tad AST tekstai pagal aktyvumo ir deskriptyvumo raišką tarpusavyje yra skirtingi. Variacijos koeficiento vertės (žr. 13 lentelę) taip pat rodo šią tendenciją. Kvartilinio pločio vertės atskleidžia, kad pagal aktyvumo ir deskriptyvumo raišką AST tekstai tarpusavyje skiriasi labiau (didesnė duomenų sklaida) nei MST ir PST (žr. 13, 20 ir 25 lenteles).

Kaip rodo 13 lentelė, asimetrijos koeficiento rodiklis parodė vidutinę neigiamą A indikatoriaus asimetriją ir vidutiniškai teigiamą indikatoriaus D asimetriją (Bulmer, 1979, 63). Toks A reikšmių pasiskirstymas reiškia, kad AST dažnesni tekstai, kuriems būdingesnė mažesnė aktyvumo raiška (mažiau veiksmožodžių). Ir kartu AST dažnesni tekstai, kuriems tipiškesnė didesnė deskriptyvumo raiška. Eksceso koeficiento vertės A ir D indikatoriams, rodo, kad duomenyse turėtų būti šiek tiek daugiau išskirčių nei tai tipiška

normaliajai kreivei. Į šį polinkį kreipiamas dėmesys, kuomet pasirenkami tyrimo metodai kitam funkcinių stilių kalbinės raiškos analizės etapui.

4.1.1.8 Tarpeiksmožodinių atstumų (VA) indikatorius savybės administraciniuose tekstuose

VA indikatorius tekste vertina vidutinį atstumą (žodžių formomis) tarp dviejų iš eilės einančių veiksmažodžių (Kubát, Matlach ir Čech, 2014, 42). VA skaičiuojamas tekstui, t. y. nekreipama dėmesio į sakinius ir ribas tarp jų. Kuo teksto sintaksinė struktūra sudėtingesnė, tuo VA vertė yra didesnė. Kaip atskleidė tyrimas, AST VA reikšmės pasiskirsčiusios nuo 3,098 (mažiausia reikšmė, žyminti mažiausią atstumą tarp 2 iš eilės einančių veiksmažodžių formų tekste) iki 622,2 (didžiausia reikšmė, žyminti didžiausią atstumą tarp 2 iš eilės einančių veiksmažodžių formų) (plg.: MST – nuo 3,233 iki 22,500 (plačiau žr. 4.2.1.8 sk.), o PST – nuo 2,652 iki 47,333 (plačiau žr. 4.3.1.8 sk.)). Sintaksės atžvilgiu paprasčiausias tekstas buvo *adm159-84.txt* – Lietuvos Respublikos Vyriausybės nutarimas dėl Lietuvos Respublikos Vyriausybės 1994 m. lapkričio 11 d. nutarimo Nr. 1123 „Dėl žyminio mokesčio tarifų bei žyminio mokesčio mokėjimo ir gražinimo tvarkos patvirtinimo“ dalinio pakeitimo. Sudėtingiausias tekstas, t. y. jame atstumas tarp 2 iš eilės einančių veiksmažodžių buvo didžiausias – *adm77-20.txt* (Lietuvos Respublikos Prezidento dekretas „Dėl Lietuvos Respublikos pilietybės atsisakymo“). Šiame tekste esama tik 3 veiksmažodžių formų (*patenkinu, atsisakyti* ir *įsigalioja*), tarp *atsisakyti* ir *įsigalioja* įsiterpia 3106 žodžiai ir žodžių formos.

Kad būtų galima įvertinti VA statistines savybes, buvo atlikta aprašomoji duomenų analizė. Jos santrauka pateikta 13-oje lentelėje. Indikatorius vidurkis (12,468) daug didesnis nei mediana (6,261). Tai rodo, kad išskirčių, galima tikėtis didelių verčių pusėj. Kitaip tariant, itin didelis atstumas tarp 2 iš eilės einančių veiksmažodžių tekste AST tekstams nėra tipiškas, o greičiau išimtis. Itin didelis, lyginant su vidurkiu, standartinio nuokrypio rodiklis (34,713) reiškia, kad VA reikšmės yra labai nutolusios nuo vidurkio (Field, Miles ir Field, 2012, 40), t. y. AST tekstams būdinga sintaksės kompleksškumo įvairovė, pagal šį požymį AST tekstai ženkliai skiriasi tarpusavyje. Variacijos koeficiento reikšmė taip pat rodo šią tendenciją. Kvartilinio pločio vertė rodo, kad pagal VA verčių sklaidos mastą AST tekstai kad ir reikšmingai skiriasi tarpusavyje, tačiau daug mažiau nei MST ir PST (žr. 13, 20 ir 25 lentelę).

Asimetrijos koeficiento vertė (8,42) rodo didelę teigiamą verčių pasiskirstymo asimetriją (Bulmer, 1979, 63). Kitaip sakant, duomenyse

didesnės vertės, žyminčios didesnius atstumus tarp 2 iš eilės einančių veiksmoždzių yra dažnesnės nei mažesnės arba AST tekstuose dažnesni yra sintaksinės struktūros požiūriu sudėtingesni tekstai. Eksceso koeficiento rodiklis (87,9) rodo, kad duomenyse turėtų būti gerokai daugiau išskirčių nei tai būdinga normaliajai kreivei ir šis netipiškų verčių mastas turės įtakos renkantis metodus kitam tyrimo etapui.

4.1.2 Indikatorių koreliacija administracinio stiliaus tekstams

Šiame poskyryje aptariama tiriamų indikatorių koreliacinė analizė, skirta statistinio ryšio tarp kintamųjų poros (šiuo atveju – indikatorių poros) įvertinimui. Ryšiui tarp kintamųjų nustatyti pasirinktas Spirmeno (Spearman) koreliacijos koeficientas (Field, Miles ir Field, 2012, 223), nes kalbinę raišką žyminčių indikatorių vertės pasižymėjo išskirtimis (Mukaka, 2012), t. y. tarp indikatorių verčių, suskaičiuotų AST tekstams, pasitaikė itin didelių arba itin mažų reikšmių, išsiskiriančių iš bendro duomenų konteksto (Čekanavičius ir Murauskas, 2001, 36).

4.1.2.1 Išskirtys administracinio stiliaus tekstyno duomenyse

Kaip galima matyti 16-oje lentelėje, visų tirtų indikatorių vertėms būdingos išskirtys, o tai reiškia, kad tarp visų indikatorių verčių buvo tokių, kurios savo didumu arba mažumu buvo netipiškos. Į tokias netipiškas indikatorių vertes bei jų mastą svarbu atsižvelgti pasirenkant tyrimo metodus, pavyzdžiui, koreliacijos koeficientą. Išskirčių intervalai indikatoriams buvo apskaičiuoti remiantis kvartilinio pločio (KP) statistika (plačiau žr. 3.1 sk.). Išskirtimis laikomos duomenų rinkinio reikšmės, kurios yra mažesnės nei $Q_1 - 1.5 KP$ arba didesnės nei $Q_3 + 1.5 KP$ (Lapinskas, 2003, 9). Šių skaičiavimų rezultatai ir pateikti 16-oje lentelėje. Formulėse Q_1 ir Q_3 – tai, atitinkamai, pirmasis kvartilis ir trečiasis kvartilis (primintina, kad kvartiliai dalina išrikiuotas pagal dydį duomenų vertes į ketvirčius).

16 lentelė. AST: indikatorių verčių išskirtys

Indikatorius	Išskirtys
VŽFI	Vertės, mažesnės nei 6,4 (rasta 17) ir didesnės nei 8,4 (rasta 56)
<i>a</i>	Vertės, didesnės nei 14 (rasta 74)
R_1	Vertės, mažesnės nei 0,63 (rasta 159)
RR_{mc}	Vertės, mažesnės nei 0,9 (rasta 166)
MATTR	Vertės, mažesnės nei 32 (rasta 208)
TK	Vertės, didesnės nei 1,4 (rasta 53)
A	Vertės, mažesnės nei 0,3 (rasta 91)
D	Vertės, didesnės nei 0,7 (rasta 91)
VA	Vertės, didesnės nei 11 (rasta 494)

Tad, kaip matoma 16-oje lentelėje, indikatorius VŽFI vertės, mažesnės nei 6,4 ir didesnės nei 8,4 patenka tarp išskirčių, t. y. vidutinis žodis ar žodžio forma, trumpesni nei 6,4 simbolių ir ilgesni nei 8,4 simbolių, laikomi netipiškais, iškreipiančiomis rezultatus, gaunamus taikant standartinius statistinius metodus. Tokių netipiškų VŽFI verčių AST tekstams iš viso rasta 73. Lygiai taip pat, indikatorius *a* išskirtys yra vertės, didesnės nei 14. Kitaip sakant, labai dažnų žodžių ir žodžių formų sritys dažninėje lentelėje, kurių įverčiai yra daugiau nei 14, yra netipiškos AST tekstams. Didesnės nei 14 vertės AST turi 74 tekstai. R_1 atveju išskirtimis tapo mažesnę žodžių ir jų formų įvairovę žyminčios vertės (mažesnės nei 0,63). Jos priklausė 159 administraciniams tekstams, kitaip sakant, šiems tekstams būdinga netipiškai maža žodžių ir žodžių formų įvairovė. Kas dėl žodžių ir jų formų koncentracijos tekstuose (šių požymi žymi indikatorius RR_{mc}), netipiškos (mažesnės nei 0,9) vertės priklausė tapo 166 tekstams, t. y. šių tekstų žodžių ir žodžių formų koncentruotumas, lyginant su visais AST tektais, buvo netipiškai mažas. Indikatorius MATTR išskirtys – vertės, mažesnės nei 32, kitaip – informacijos srautas, kurio įvertis mažesnis nei 32 yra netipiškas AST tekstams. AST rasti 208 tekstai su tokiomis vertėmis.

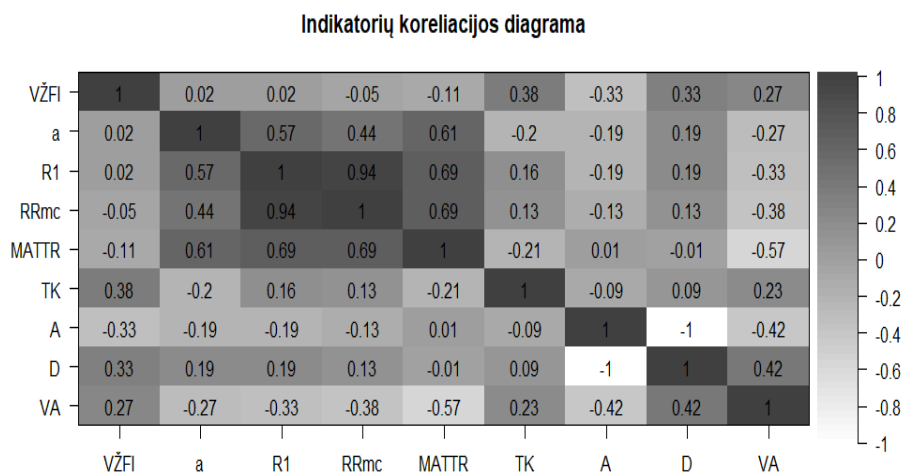
TK išskirtimis laikomos vertės, didesnės nei 1,4, o tokių tekstų rasta 53. Kitaip sakant, itin didelė teminė koncentracija yra netipiška AST tekstams, tad taikant standartinius statistinius metodus tokios netipiškos vertės iškreiptų tyrimo rezultatus. Tarp teksto aktyvumą / dinamiškumą žyminčio indikatorius A verčių netipiški atvejai yra tekstai, kurių aktyvumo raiška itin maža (indikatorius A vertė mažesnė nei 0,3), o teksto deskriptyvumą žyminčio indikatorius D atžvilgiu – tekstai, kurių deskriptyvumo raiška itin didelė (indikatorius D vertė mažesnė nei 0,3). Kadangi šie indikatoriai yra vienas kitam priešingi, išskirtimis (netipiškais atvejais) tapo tas pats 91 tekstas, kurių aktyvumo raiška labai maža, o deskriptyvumo – labai didelė.

Galiausiai AST tekstai, kurių VA indikatoriaus buvo vertės didesnės nei 11 taip pat priklauso netipiškiems atvejams. Tekstų, kurių vidutinis atstumas tarp 2 iš eilės einančių veiksmažodžių yra daugiau nei 11 rasta 494.

Kaip galima matyti šiame poskyryje, AST tekstams suskaičiuotų visų indikatorių vertėms būdingos išskirtys – netipiškai didelės arba netipiškai mažos vertės. Ši situacija lėmė, kad koreliacinei analizei buvo pasirinktas ne įprastas Pirsono koreliacijos koeficientas, bet ranginis Spirmeno, kuris yra tinkamas, kai duomenims būdingos išskirtys (Mukaka, 2012).

4.1.2.2 Koreliacinės analizės rezultatai AST

AST indikatorių koreliacinės analizės rezultatai susistemintai pateikti 5-ame paveiksle. „Karščio žemėlapis“ diagramos (Gu, Eils ir Schlesner, 2016) kvadratėliuose pateiktos koreliacijos koeficiento vertės žymi ryšio tarp kintamųjų stiprumą kintamųjų, t. y. indikatorių, poroms. Kuo šviesesne spalva nuspalvintas kvadratėlis, tuo stipresnis neigiamas ryšys sieja 2 indikatorius, t. y. vieno jų reikšmėms didėjant, kito – mažėja ir atvirkščiai. Lygiai taip pat, kuo tamsesne spalva nuspalvintas kvadratėlis, tuo stipresnis teigiamas ryšys sieja 2 indikatorius, t. y. vieno jų reikšmėms didėjant, kito taip pat didėja, o mažėjant – mažėja.



5 pav. Koreliacinės analizės diagrama AST

5-ame paveiksle pateiktoje diagramoje galima matyti, kad labai stipri koreliacija nustatyta 2 indikatorių poroms – RR_{mc} ir R_1 bei A ir D ¹⁰⁸. Vidutiniškai stiprus ryšys sieja 8 indikatorių poras: R_1 ir a ; RR_{mc} ir a ; $MATTR$ ir a ; $MATTR$ ir R_1 ; $MATTR$ ir RR_{mc} ; $MATTR$ ir VA ; A ir VA ; D ir VA . O silpna ir labai silpna koreliacija rasta tarp 10-ies indikatorių porų: TK ir $V\check{Z}FI$, A ir $V\check{Z}FI$, D ir $V\check{Z}FI$, VA ir $V\check{Z}FI$, TK ir a , VA ir a , VA ir R_1 , VA ir RR_{mc} , TK ir $MATTR$, VA ir TK . 17-oje lentelėje pateikiami ryšių tarp indikatorių porų statistinio reikšmingumo įverčiai.

17 lentelė. AST: koreliacijos tarp indikatorių porų statistinis reikšmingumas (p -vertės)

	VŽFI	a	R_1	RR_{mc}	MATTR	TK	A	D	VA
VŽFI		0,20 0	0,16 8	0,001	0,000	0,00 0	0,00 0	0,00 0	0,0 00
a	0,200		0,00 0	0,000	0,000	0,00 0	0,00 0	0,00 0	0,0 00
R_1	0,168	0,00 0		0,000	0,000	0,00 0	0,00 0	0,00 0	0,0 00
RR_{mc}	0,001	0,00 0	0,00 0		0,000	0,00 0	0,00 0	0,00 0	0,0 00
MATTR	0,000	0,00 0	0,00 0	0,000		0,00 0	0,44 5	0,44 5	0,0 00
TK	0,000	0,00 0	0,00 0	0,000	0,000		0,00 0	0,00 0	0,0 00
A	0,000	0,00 0	0,00 0	0,000	0,445	0,00 0		0,00 0	0,0 00
D	0,000	0,00 0	0,00 0	0,000	0,445	0,00 0	0,00 0		0,0 00
VA	0,000	0,00 0	0,00 0	0,000	0,000	0,00 0	0,00 0	0,00 0	

Kaip galima matyti 17 lentelėje, koreliacija buvo statistiškai reikšminga ($p \leq 0,05$) visoms indikatorių poroms, išskyrus 4: A ir RR_{mc} ; D ir RR_{mc} ; VA ir R_1 ; VA ir RR_{mc} . Šioms indikatorių poroms nustatytas ryšys buvo statistiškai nereikšmingas, t. y. $p \geq 0,05$. Poskyriuose 4.1.2.2.1-4.1.2.2.3 pateikiami detalūs koreliacinės analizės AST rezultatai.

¹⁰⁸ Nors A ir D yra aiškiai priklausomi kintamieji, tad vieno jų būtų galima atsisakyti, tačiau, siekiant susidaryti bendrą vaizdą ir sekant čekų mokslininkų darbais (pavyzdžiui, Kubát (2015)), šiame tyrimo etape pasiekiami abu indikatoriai. Vis dėlto, 5 sk. FS tarpusavio palyginimui taikant neparimetrinius išvadų statistikos testus, iš šių dviejų indikatorių pasilieka tik indikatorius A .

4.1.2.2.1 Labai stipri koreliacija tarp indikatorių

Duomenų analizė atskleidė, kad labai stiprios statistiškai reikšmingos koreliacijos (Čekanavičius ir Murauskas, 2001, 126) esama tarp indikatorių RR_{mc} ir R_1 (0,94; žymi labai stiprią teigiamą koreliaciją) bei A ir D (-1; žymi labai stiprią neigiamą koreliaciją). Indikatorius RR_{mc} leidžia įvertinti AST tekstų žodžių ir žodžių formų koncentruotumą, o R_1 – (retesnių) žodžių ir žodžių formų įvairovę. Tad kuo tekste didesnis žodžių ir žodžių formų koncentruotumas (daugiau itin dažnai pasikartojančių žodžių ir žodžių formų), tuo jam būdingesnė ir didesnė (retesnių) žodžių ir žodžių formų įvairovė. RR_{mc} apima visus tekstą sudarančius žodžius, o R_1 – tik dalį, t. y. retesnius, dažninėje lentelėje esančius žemiau h-taško (plačiau žr. 3.1.2.1 sk.). Tad iš esmės visai įmanoma situacija, kuomet turime tekstą, kuriame pora žodžių pasikartoja itin daug kartų, o visi likę – vos po porą ar tik vieną. Tokiu būdu ir gauname tiek didelį / mažą žodžių ir žodžių formų koncentruotumą, tiek didelę / mažą (retesnių) žodžių ir žodžių formų įvairovę.

Indikatoriai A ir D žymi teksto „toną“: A – teksto aktyvumą, o D – deskriptyvumą. A ir D yra vienas kitam priešingi. Aktyvumas apibūdinamas veiksmažodžių formų ir būdvardžių formų santykiu, o deskriptyvumas – būdvardžių formų ir veiksmažodžių formų santykiu. Tad abu šiuos indikatorius, kaip ir buvo galima tikėtis, sieja itin stiprus neigiamas statistiškai reikšmingas ryšys, kuris rodo, kad tekste stiprėjant aktyvumo raiškai, deskriptyvumo raiška mažėja ir atvirkščiai.

4.1.2.2.2 Vidutiniškai stipri koreliacija tarp indikatorių

Vidutinio stiprumo (Čekanavičius ir Murauskas, 2001, 126) statistiškai reikšminga koreliacija nustatyta 8 indikatorių poroms: R_1 ir a ; RR_{mc} ir a ; MATTR ir a ; MATTR ir R_1 ; MATTR ir RR_{mc} ; MATTR ir VA; A ir VA; D ir VA (žr. 5 pav.). Toliau plačiau apžvelgiami vidutinio stiprumo ryšiai, nustatyti šių indikatorių poroms.

Kaip minėta, R_1 žymi retesnių, t. y. priklausančių pagrindinėms kalbos dalims, žodžių ir žodžių formų įvairovę (turingumą). Indikatorius a žymi labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų srities dydį dažninėje lentelėje. Tad teigiamas ryšys tarp šių indikatorių rodo, kad kuo tekste mažiau (retesnių) žodžių ir žodžių formų, tuo jam būdingesnė didesnė labai dažnų žodžių sritis ir atvirkščiai.

Primintina, kad RR_{mc} šios disertacijos ribose žymi žodžių ir žodžių formų koncentruotumą tekste. Tad teigiamas ryšys tarp indikatorių RR_{mc} ir a rodytų, kad kuo didesnis žodžių ir jų formų koncentruotumas (kai vos keletas žodžių ir žodžių formų pasikartoja itin dažnai), tuo tekstui būdingesnė mažesnė labai dažnų žodžių sritis ir atvirkščiai.

MATTR yra informacijos srauto/temos išdėstymo įvertis (Popescu, 2009, 233—235). Teigiamas ryšys tarp MATTR ir a rodo, kad kuo didesnis/mažesnis informacijos srautas, t. y. kuo daugiau tekste skirtingų žodžių ir jų formų, tuo mažesnė/didesnė labai dažnų žodžių ir žodžių formų sritis būdingesnė tekstui. MATTR ir R_1 indikatorių porą siejantis teigiamas ryšys rodo, kad kuo tekstas pasižymi didesniu/mažesniu informacijos srautu (kitaip – didesniu skaičiumi skirtingų žodžių ir jų formų), tuo didesnė jame (retesnių) žodžių ir žodžių formų įvairovė. Tarp indikatorių MATTR ir RR_{mc} taip pat nustatytas vidutinio stiprumo teigiamas ryšys. Tai rodo, kad kuo didesnis/mažesnis informacijos srautas (didesnės/mažesnės MATTR vertės), tuo didesnė/mažesnė tekstą sudarančių žodžių ir jų formų koncentracija.

VA žymi vidutinį atstumą tarp 2 iš eilės tekste einančių veiksmažodžių. Jį taip pat galima laikyti tam tikru teksto sintaksinės struktūros kompleksiško įverčiu (Kubát, 2016, 70—71). Tad vidutinio stiprumo neigiamas ryšys tarp VA ir MATTR rodo, kad kuo paprastesnė teksto sintaksinė struktūra, t. y. kuo mažesnis jame atstumas tarp 2 iš eilės einančių veiksmažodžių, tuo jam būdingesnis didesnis informacijos srautas ir atvirkščiai – kuo teksto sintaksinė struktūra sudėtingesnė, tuo AST tekste mažesnis informacijos srautas (kitaip – tuo mažiau tekste skirtingų žodžių ir jų formų).

Indikatorius A žymi vieną teksto „tono“ ypatybių – jo aktyvumą, išreikštą veiksmažodžių ir būdvardžių santykiu tekste. Indikatorius A ir Va sieja neigiama koreliacija. Kitai tariant, kuo didesne A raiška pasižymi tekstas (t. y. kuo daugiau jame veiksmažodžių), tuo paprastesnė jo sintaksinė struktūra (t. y. tuo vidutiniškai mažesnis tekste atstumas tarp 2 iš eilės einančių veiksmažodžių). Taip pat ir atvirkščiai: kuo mažesnė A raiška AST tekste, tuo jo sintaksinė struktūra yra sudėtingesnė. Indikatorius D žymi teksto deskriptyvumą tekste per būdvardžių santykį su veiksmažodžiais, t. y. D yra atvirkščias indikatorius A. Tad teigiama koreliacija tarp D ir VA rodo, kad kuo didesnė/mažesnė pirmojo indikatoriaus raiška, t. y. kuo daugiau/mažiau administracinio stiliaus tekste būdvardžių, lyginant su veiksmažodžiais, tuo teksto sintaksinė struktūra yra sudėtingesnė/paprastesnė, kitai – tuo didesnis/mažesnis atstumas tekste tarp 2 iš eilės einančių veiksmažodžių formų.

4.1.2.2.3 Silpna ir labai silpna koreliacija tarp indikatorių

Silpna koreliacija (koreliacijos koeficiento reikšmės $[-0,4;-0,2]$ arba $(0,2; 0,4]$ (Čekanavičius ir Murauskas, 2001, 126)) nustatyta 10 indikatorių porų: TK ir VŽFI, A ir VŽFI, D ir VŽFI, VA ir VŽFI, TK ir a , VA ir a , VA ir R_1 , VA ir RR_{mc} , TK ir MATTR, VA ir TK (5 pav.). Tai rodo, kad šių indikatorių poros turi nedidelių sąsajų – teigiamų, kai abiejų indikatorių porose reikšmės didėja/mažėja (TK ir VŽFI, D ir VŽFI, VA ir VŽFI, VA ir TK), arba neigiamų, kai vieno poros indikatorius reikšmės didėja, o kito – mažėja arba atvirkščiai (A ir VŽFI, TK ir a , VA ir a , VA ir R_1 , VA ir RR_{mc} , VA ir MATTR). Labai silpna koreliacija (Čekanavičius ir Murauskas, 2001, 126) nustatyta 14 indikatorių porų (5 pav.). Iš jų 8 poras sieja neigiamas ryšys (RR_{mc} ir VŽFI, MATTR ir VŽFI, A ir a , A ir R_1 , A ir RR_{mc} , a ir TK, MATTR ir TK, A ir TK), o 6 – teigiamas (D ir a , TK ir R_1 , D ir R_1 , TK ir RR_{mc} , D ir RR_{mc} , D ir TK).

Galiausiai koreliacija buvo statistiškai nereikšminga, t. y. $p \geq 0,05$, 4 indikatorių poros: A ir RR_{mc} ; D ir RR_{mc} ; VA ir R_1 ; VA ir RR_{mc} (žr. 17 lentelę). Kitaip sakant, šioms poroms vieno poroje esančio indikatorius reikšmių kitimas neturi statistiškai reikšmingo poveikio kito toje pačioje poroje esančio indikatorius reikšmių kitimui.

4.1.3 Pagrindinių komponentių analizės (PKA) rezultatai AST

4.1.3.1 Indikatorių svarba reprezentuojant administracinio stiliaus kalbinę raišką

Siekiant išskirti, kurie tiriami indikatorius bei jų kombinacija apibūdina administracinio FS kalbinę raišką geriausiai, buvo atlikta pagrindinių komponentių analizė (PKA). Šis metodas leido įvertinti kiekvieno iš 9 indikatorių, naudotų tyrime (VŽFI, a , R_1 , RR_{mc} , MATTR, TK, A, D ir VA), indėlį ir svarbą apibūdinant šį FS bei išskirti geriausiai administracinį stilių apibūdinančių indikatorių kombinaciją. 18-oje lentelėje pateikiama administracinio FS kalbinę raišką reprezentuojančių indikatorių PKA rezultatų santrauka (lentelėje „PK1“, ..., „PK9“ žymimos pagrindinės komponentės). Kaip jau minėta, atliekant analizę duomenų „debesis“, t. y. indikatorių vertės, sukiojamos, ieškant krypties, kurioje duomenų dispersija, t. y. sklaida, yra didžiausia. Toji kryptis vadinama pirmąja pagrindine komponente. Iš esmės tai yra normalizuota linijinė kintamųjų (šiuo atveju – 9-ių kalbinę raišką žyminčių indikatorių) iš pirminio, „nepasukto“ duomenyno kombinacija.

Kaip galima matyti iš 18-oje lentelėje pateiktos informacijos, pirmos trys pagrindinės komponentės paaiškina didžiąją dalį visos duomenų dispersijos. Pirmoji pagrindinė komponentė PK1 apibūdina 35,8 proc. duomenų variacijos, antroji pagrindinė komponentė PK2 – 25,4 proc. duomenų variacijos, o trečioji – 13,8 proc. duomenų variacijos (žr. 18-os lentelės eilutę „Variacijos proporcija, %“). Kartu jos paaiškina 75,00 proc. visos duomenų variacijos ($PK1 + PK2 + PK3 = 35,8 + 25,4 + 13,8 = 75,00$; žr. 18-os lentelės eilutę „Sukauptinė variacijos proporcija, %“). Likusios 6 komponentės kiekviena atskirai paaiškina vis mažesnę duomenų variacijos dalį (žr. 18-os lentelės eilutę „Variacijos proporcija, %“). O kartu jos paaiškina tik 25 proc. kintamųjų reikšmių ($PK4 + PK5 + PK6 + PK7 + PK8 + PK9 = 10,55 + 6,59 + 5,25 + 1,97 + 0,64 + 0,00 = 25$; žr. 18-os lentelės eilutę „Variacijos proporcija, %“). Visų kintamųjų (kalbinę raišką žyminčių indikatorių) reikšmių atvaizdavimui pakanka 7 komponentių (sudėjus pirmas 7 komponentes, jos visos kartu paaiškina 99,36 proc. visos reikšmių variacijos). Tad, kaip įprasta atliekant PKA, tolesnei analizei pasiliekomos tik didžiausią kintamųjų reikšmių variacijos dalį paaiškinančios komponentės PK1, PK2 ir PK3. Kitaip sakant, šioms komponentėms priklauso didžioji indikatorių verčių dalis ir jose esančios indikatorių verčių kombinacijos leidžia išskirti indikatorius, kurių, savo ruožtu, verčių dalis tose komponentėse yra didžiausia.

18 lentelė. PKA rezultatų santrauka AST: duomenų variacijos proporcijos, kurias paaiškina komponentės

	PK1	PK2	PK3	PK4	PK5	PK6	PK7	PK8	PK9
Tikrinių vektorių reikšmės	3,22	2,29	1,24	0,95	0,59	0,47	0,18	0,06	0,00
Variacijos proporcija (%)	35,80	25,40	13,80	10,55	6,59	5,25	1,97	0,64	0,00
Sukauptinė variacijos proporcija (%)	35,80	61,20	75,00	85,55	92,14	97,39	99,36	100,00	100,00

Paprastai tolesnei analizei ir pasirenkamos tos pagrindinės komponentės (jos iš esmės yra kintamųjų reikšmių kombinacijos), kurios paaiškina daugiau duomenų variacijos nei atskiri kintamieji pradiniam duomenų rinkinyje, t. y. tik tos, kurių tikrinių vektorių reikšmės yra didesnės nei 1 (Kassambara, 2017,

18). Vienos komponentės PK1 nepakanka, kadangi ji apibūdina tik truputį daugiau nei trečdalį (35,80 %) visos AST kalbinės raiškos. Tad, kaip jau minėta, tolesnei analizei pasiliekomos pagrindinės komponentės PK1, PK2 ir PK3, nes būtent jų tikrinių vektorių reikšmės yra didesnės nei 1 (18 lentelė). Remiantis PKA rezultatais, pateiktais 18-oje lentelėje, toliau buvo analizuojama, kurie indikatoriai tiriant administracinio funkcinio stiliaus kalbinę raišką reprezentuoja didžiausią dalį kintamųjų reikšmių, t. y. yra svarbiausi arba, kitaip sakant, geriausiai reprezentuoja šio stiliaus raišką. Be to, siekta nustatyti kiekvieno indikatoriaus indėlį charakterizuojant šį FS. Konkrečių indikatorių svarba įvertinta pagal kampo tarp indikatorių verčių koordinacių didumą – kvadratinės kosinuso vertes (\cos^2) (Husson, Josse, Le ir Mazet, 2018). Kuo didesnė ši vertė, tuo indikatorius konkrečioje komponentėje yra svarbesnis, t. y. komponentėje tuo geriau jis reprezentuojamas. Kiekvieno indikatoriaus indėlis konkrečioje pagrindinėje komponentėje įvertinamas procentais, t. y. skaičiuojama, kokią dalį tam tikrai komponentei priklausančių verčių išreiškia atskiri indikatoriai. Kuo ši procentinė išraiška didesnė, tuo indikatoriaus indėlis konkrečioje komponentėje yra didesnis.

19 lentelė. PKA rezultatų santrauka AST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis

	PK1		PK2		PK3	
	Indėlis (%)	Svarba (\cos^2)	Indėlis (%)	Svarba (\cos^2)	Indėlis(%)	Svarba (\cos^2)
VŽFI	0,114	0,004	16,183	0,370	18,206	0,226
<i>a</i>	12,351	0,398	0,165	0,004	3,262	0,041
R_1	27,396	0,883	0,173	0,004	3,049	0,038
RR_{mc}	25,636	0,826	0,111	0,003	3,301	0,041
MATTR	23,760	0,766	2,261	0,052	0,000	0,000
TK	0,119	0,004	5,554	0,127	54,782	0,681
A	0,629	0,020	37,532	0,858	8,065	0,100
D	0,629	0,020	37,532	0,858	8,065	0,100
VA	9,367	0,302	0,489	0,011	1,271	0,016

Kaip galima matyti 19-oje lentelėje, PK1 svarbiausi, t. y. geriausiai reprezentuojami, yra R_1 , RR_{mc} ir MATTR indikatoriai (jų \cos^2 vertės yra didžiausios). Jų indėlis šioje komponentėje taip pat yra didžiausias (žr. 19-ą lentelę). Kita vertus, PK1 mažiausia svarbūs (kartu ir mažiausiai prisidėję

reprezentuojant kintamųjų reikšmes) yra indikatoriai VŽFI ir TK. Tad PK1, kuri reprezentuoja 35,8 proc. visų kintamųjų (t. y. visų 9 indikatorių) reikšmių (18 lentelė), svarbiausi yra 3 su dažnine teksto struktūra susiję indikatoriai (pagal svarbą): R_1 , RR_{mc} ir MATTR. Kartu šie indikatoriai apibūdina beveik 77 proc. PK1 reikšmių (R_1 (27,396 proc.) + RR_{mc} (25,636 proc.) + MATTR (23,760) = 76,79 proc.; žr. 19-ą lentelę). Taigi galima sakyti, kad PK1 šie indikatoriai – R_1 , RR_{mc} ir MATTR – geriausiai reprezentuoja disertacijoje analizuotus administracinio stiliaus tekstus. Ši indikatorių kombinacija paaikškina daugiau nei ketvirtadalį visos administracinio FS raiškos, t. y. 27,5 proc. (PK1 apibūdina 35,8 proc. visos administracinio stiliaus, o R_1 , RR_{mc} ir MATTR kombinacija – 76,79 proc. šio stiliaus PK1 raiškos).

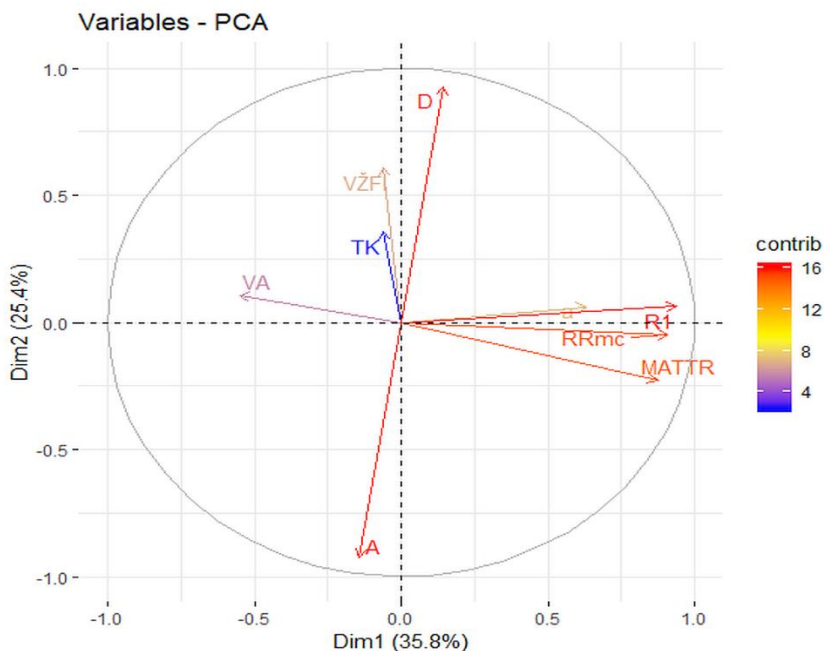
Komponentė PK2 reprezentuoja 25,4 proc. visų 9 tirtų indikatorių reikšmių (žr. 18 lentelę). Pagal svarbą ir indėlį šioje komponentėje išsiskiria indikatoriai A ir D, turintys didžiausias \cos^2 vertes (žr. 19 lentelę). Abiejų indėlis yra po 37,532 proc., tad abu kartu šie indikatoriai reprezentuoja net 75,06 proc. visų PK2 priklausančių kintamųjų reikšmių. Indikatoriai A ir D naudojami kaip vienas iš būdų nusakyti teksto „toną“: A – aktyvumą per veiksmoždžių santykį su būdvardžiais, o D – deskriptyvumą per būdvardžių santykį su veiksmoždžiais. Mažiausiai svarbūs ir kartu mažiausia prisidėję prie kintamųjų reikšmių reprezentavimo PK2 yra indikatoriai RR_{mc} , R_1 ir a (19 lentelė). Taigi PK2 kontekste disertacijoje analizuoti AST tekstai geriausiai apibūdinami per aktyvumo ir deskriptyvumo dimensijas. Šie indikatoriai kartu apibūdina 19,07 proc. visos AST raiškos (PK2 apibūdina 25,4 proc. visos administracinio stiliaus, o A ir D kombinacija – 76,79 proc. šio stiliaus PK2 raiškos).

Paskutinėje analizuotoje pagrindinėje komponentėje PK3 išsiskiria vienas indikatorius – TK, apibūdinantis teksto teminę koncentraciją. Šio indikatoriaus svarba ir indėlis smarkiai skiriasi nuo kitų šios komponentės indikatorių: TK \cos^2 vertė yra 0,681 ir šis indikatorius reprezentuoja net 54,782 proc. visų PK3 kintamųjų reikšmių. Tad PK3 atžvilgiu disertacijoje analizuoti administracinio stiliaus tekstai geriausiai apibūdinami per jų teminės koncentracijos laipsnį, t. y. pagal tai, kaip labai tekstai yra sukoncentruoti į savo temą. Šis indikatorius nusako 7,6 proc. visos administracinio stiliaus kalbinės raiškos (PK3 apibūdina 13,8 proc. visos AST indikatorių verčių dispersijos, o TK – 54,782 proc. šio stiliaus PK3 duomenų sklaidos).

4.1.3.2 Indikatorių indėlis reprezentuojant administracinio stiliaus kalbinę raišką

Ištyrus atskirų indikatorių svarbą charakterizuojant AST kalbinę raišką, atlikta ir indikatorių indėlio apibūdinant šio stiliaus kalbinę raišką analizė, kurios rezultatai pateikti 6-ame ir 7-ame paveiksluose. Kuo indikatorius reprezentuojančių rodyklių yra arčiau raudonos spalvos spektro, tuo jų indėlis apibūdinant administracinio stiliaus kalbinę raišką yra didesnis. Ir atvirkščiai: kuo indikatorius reprezentuojančių rodyklių spalva yra arčiau violetinės spalvos spektro, tuo jų indėlis apibūdinant šio FS kalbinę raišką yra mažesnis. Taip pat, kuo šių rodyklių smaigaliai yra arčiau paveikslė pavaizduoto apskritimo perimetro, tuo indikatorių indėlis taip pat didesnis, o kuo jie yra arčiau paveikslė pavaizduoto apskritimo centro, tuo indikatorių indėlis yra mažesnis.

Taigi 6-ame paveiksle pavaizduoti atskirų indikatorių indėliai apibūdinant administracinio FS tekstyno tekstų pirmųjų dviejų pagrindinių komponentų PK1 ir PK2 kontekste. X (horizontalioje) ašyje atvaizduota PK1 (ši komponentė reprezentuoja didžiausią dalį indikatorių reikšmių variacijos, t. y. 35,8 proc.), o Y (vertikaliąje) ašyje – PK2 (ši komponentė reprezentuoja didžiausią dalį indikatorių reikšmių variacijos, likusios po PK1 t. y. 25,4 proc.). Kaip ir indikatorių svarbos atveju, taip ir šiuo galioja ta pati taisyklė: kuo kintamąjį (indikatorių) reprezentuojanti rodyklė yra arčiau paveikslė pavaizduoto apskritimo perimetro, tuo indikatoriaus indėlis yra didesnis, o kuo ši rodyklė yra arčiau paveikslė pavaizduoto apskritimo centro, tuo indikatoriaus indėlis yra mažesnis. Ir kartu kuo minėtų rodyklių spalva arčiau raudonos spalvos spektro, tuo indikatoriaus indėlis taip pat yra didesnis, o kuo arčiau mėlynos – tuo indikatoriaus indėlis yra mažesnis.

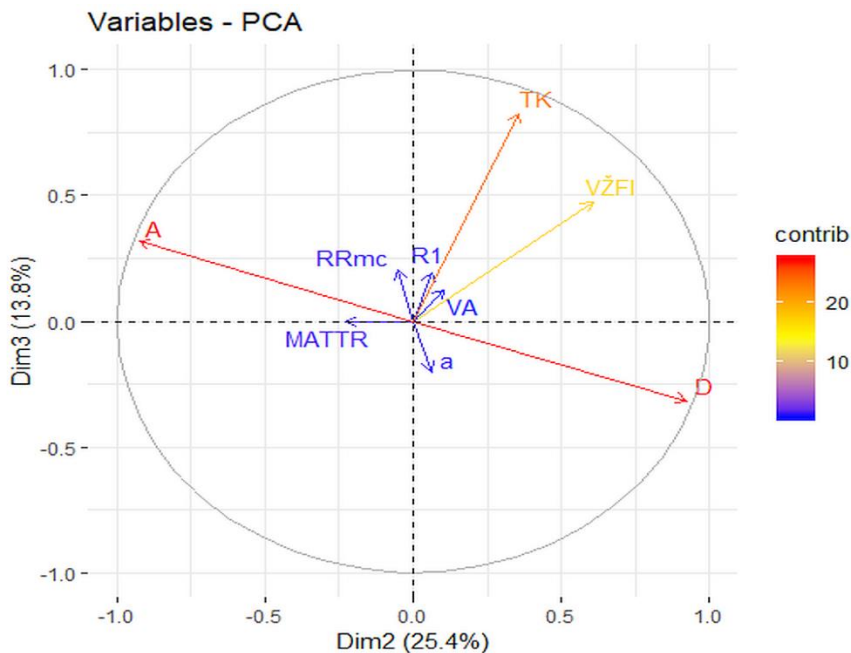


6 pav. AST: indikatorių indėlis (PK1 ir PK2)

Tad galima matyti, kad didžiausias indėlis apibūdinant administracinio stiliaus tekstus PK1 ir PK2 kontekste priklauso indikatoriams R_1 (žymi (retesnių) žodžių ir žodžių formų įvairovę), RR_{mc} (žymi žodžių ir žodžių formų koncentruotumą), MATTR (informacijos srauto įvertis), A (teksto aktyvumo indikatorius) ir D (teksto deskriptyvumo indikatorius) (6 pav.) – juos vaizduojančių rodyklių spalva yra arčiausiai raudonos spalvos spektro, o tų pačių rodyklių smaigaliai baigiasi arčiausiai paveiksle pavaizduoto apskritimo perimetro. Kitaip sakant, PK1 ir PK2, reprezentuojančiose daugiau nei pusę AST kalbinę raišką apibūdinančių kintamųjų reikšmių variacijos (žr. 19-ą lentelę), (retesnių) žodžių ir jų formų įvairovė (R_1), žodžių ir jų formų koncentruotumas (RR_{mc}), MATTR (informacijos srauto arba temos išdėstymo tekste įvertis) bei tekstų „tonas“ (aktyvumo-deskriptyvumo dimensijos, t. y. indikatoriai A ir D) labiausiai prisideda charakterizuojant šį funkcinį stilų.

Indikatorių svarba pagal jų indėlį reprezentuojant AST pagrindinėje komponentėje PK3 pavaizduota 7-ame paveiksle. Kaip ir indikatorių svarbos pagal \cos^2 atveju, taip ir analizuojant indėlių reprezentuojant AST, siekiant kaip galima tiksliau atvaizduoti indėlių reikšmes antrąja ašimi dvimatei diagramai pasirenkama antroji pagrindinė komponentė PK2, jau aptarta aukščiau. Tad 7-ame paveiksle X (horizontalioje) ašyje pavaizduota PK2 (nes ji reprezentuoja didesnę procentą visų kintamųjų reikšmių),

apibūdinanti 25,4 proc. visų kintamųjų (indikatorių) reikšmių, o Y (vertikaliajoje) ašyje pateikiama PK3, apibūdinanti 13,8 proc. visų kintamųjų reikšmių.



7 pav. AST: indikatorių indėlis (PK2 ir PK3)

Taigi kaip matyti 7-ame paveiksle, PK2 ir PK3 kontekste indikatoriai A, D ir TK daugiausiai prisideda apibūdinant administracinį funkcinį stilių – juos reprezentuojančių rodyklių spalva yra raudoname spalvos spektre arba arti jo, o jų smaigaliai baigiasi arčiausiai paveiksle pavaizduoto apskritimo perimetro. Indikatoriai A ir D apibūdina teksto „toną“ per aktyvumo-deskriptyvumo dimensijas. Jų indėlis administracinio stiliaus tekstų apibūdinimui aptartas analizuojant 6-ą paveikslą. Indikatorius TK įvertina teksto teminę koncentraciją, t. y. kaip tekstas yra sukonzentruotas į savo temą. Mažiausiai PK2 ir PK3 kontekste apibūdinant publicistinį funkcinį stilių prisideda RR_{mc}, MATTR, R₁, a ir VA – šiuos indikatorius žyminčių rodyklių spalvos yra violetiniame-mėlyname spektre, o jų smaigaliai yra arčiausiai paveiksle pavaizduoto apskritimo centro. Tad PK2 ir PK3 atžvilgiu administracinis funkcinis stilius geriausiai apibūdinamas per teksto „toną“ (aktyvumo-aprašomumo dimensija ir atitinkamai indikatoriai A ir D) bei teminę koncentraciją (TK). Kadangi analizuojamos tik tam tikroms pagrindinėms komponentėms (pavyzdžiui, PK2 ir PK3), indikatorių svarba skiriasi. Taip yra todėl, kad pagrindinėms komponentėms priklauso skirtingos indikatorių

verčių kombinacijos. Kitaip sakant, vienoje pagrindinėje komponentėje vieno indikatorius verčių gali būti daugiau (toje komponentėje jie bus svarbesni apibūdinant konkretaus FS raišką), kitoje – kitų.

4.2 Mokslinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė

4.2.1 Aprašomoji (pirminė) mokslinių tekstų požymių analizė

Siekiant apibendrinti mokslinio stiliaus tekstų kalbinės raiškos požymių statistines savybes, atlikta aprašomoji duomenų analizė (Casas, 2016). Bazinių statistinių charakteristikų santrauka kiekvienam tiriamam indikatoriumi (indikatorių vertės buvo suskaičiuotos kiekvienam MST tekstui) pateikiama 20-oje lentelėje. Detalūs rezultatai pateikiami 4.2.1.1-4.2.1.8 sk.

20 lentelė. MST: bazinių statistinių charakteristikų santrauka

Indikatorius	Vidurkis	Mediana	Standartinis nuokrypis	Variacijos koef.	Kvartilinis plotis	Asimetrijos koef.	Eksceso koef.
VŽFI	6,986	7,051	0,543	0,078	0,755	0,186	7,3
<i>a</i>	10,211	10,013	2,015	0,197	2,399	1,147	7,6
R_1	0,8111	0,8170	0,051	0,063	0,070	-0,680	3,9
RR_{mc}	0,9526	0,9537	0,010	0,011	0,011	-1,042	5,6
MATTR	43,76	44,08	2,026	0,046	2,780	-0,823	3,7
TK	0,1558 8	0,1188 9	0,145	0,933	0,186	1,581	6,3
A	0,6204	0,6199	0,086	0,139	0,116	-0,062	2,9
D	0,3796	0,3801	0,086	0,228	0,116	0,062	2,9
VA	5,314	5,208	1,037	0,195	1,051	4,899	76,1

Vidurkis ir mediana žymi „tipiškas“ kiekvieno indikatorius vertes, standartinis nuokrypis, variacijos koeficientas ir kvartilinis plotis leidžia įvertinti indikatorius verčių sklaidą, t. y. kaip labai moksliniai tekstai pagal

konkretų indikatorių skiriasi tarpusavyje. Asimetrijos koeficientas leidžia įvertinti, kokios konkretaus indikatoriaus vertės yra dažnesnės – didesnės ar mažesnės. Kitaip tariant, sudaromos prielaidos įvertinti tam tikro indikatoriaus žymimo kalbinės raiškos požymio stiprumą. Eksceso koeficientas yra skirtas išskirčių duomenyse įvertinimui bei tam tikrų netipiškų atvejų pagal tam tikro indikatoriaus raišką masto nustatymui, o tai turi įtakos tolesnių tyrimo metodų pasirinkimui.

4.2.1.1 Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus charakteristikos mokslinio stiliaus tekstuose

MST tekstams VŽFI vertės pasiskirsčiusios intervale nuo 5,208 iki 11,536. Kitaip tariant, vidutinis žodžio formos ilgis atskiriems MST tekstams yra nuo 5,208 iki 11,536 simbolių (plg.: AST – nuo 5,373 iki 9,209 simbolių (plačiau žr. 4.1.1.1 sk.)), o PST – nuo 5,013 iki 7,945 simbolių (plačiau žr. 4.3.1.1 sk.)). Verčių vidurkis yra mažesnis nei mediana (žr. 20 lentelę), o tai rodo, kad išskirčių, VŽFI vertės išdėsčius pagal dydį, galima tikėtis mažų verčių pusėj. Kitaip sakant – vidutiniškai itin trumpi žodžiai ir žodžių formos yra netipiški MST.

Kaip galima matyti 20 lentelėje, standartinio nuokrypio vertė, lyginant su vidurkiu yra vidutiniškai didelė, tad tai rodo, kad VŽFI vertės yra vidutiniškai artimos vidurkiui, t. y. MST vidutinis žodžių ir žodžių formų ilgio raiška turėtų linkti didesnės įvairovės link. Variacijos koeficientas taip pat rodo, kad MST tekstai pagal VŽFI tarpusavyje skiriasi vidutiniškai, nors ir kiek labiau nei AST ir PST tekstai (žr. 13 ir 25 lenteles). Pagal kvartilinio pločio vertę (20 lentelė) galima sakyti, kad MST tekstai pagal vidutinio žodžio ilgio parametą tarpusavyje skiriasi labiausiai, PST – mažiau nei MST (žr. 25 lentelę), o AST tekstai šiuo atžvilgiu skiriasi mažiausiai (žr. 13 lentelę).

Asimetrijos koeficiento vertė nusako, kad didesnės VŽFI vertės MST yra būdingesnės (asimetrijos koeficiento vertė rodo nedidelę teigiamą asimetriją, žr. 20 lentelę), nors ši tendencija yra silpnesnė nei AST tekstų (žr. 13 lentelę). Kitaip sakant, MST būdingesni vidutiniškai ilgesni žodžiai ir žodžių formos, nors šis polinkis stipresnis AST tekstams. Eksceso koeficiento vertė atskleidžia, kad MST būdingos VŽFI verčių išskirtys (žr. 20 lentelę), t. y. netipiški vidutinio žodžių ir žodžių formų ilgio atvejai. Į šią tendenciją atsižvelgiama renkantis tolesnius funkcinių stilių kalbinės raiškos analizės metodus.

4.2.1.2 Indikatoriaus a charakteristikos MST

Indikatoriaus a vertės MST tekstams pasiskirsčiusios nuo 5,450 iki 25,429 (plg. AST – 1,576 iki 39,50 (plačiau žr. 4.1.1.2 sk.), o PST – nuo 3,59 iki 43,22 (plačiau žr. 4.3.1.2 sk.)). Šios vertės rodo labai dažnų (paprastai – funkcinių, nors nebūtinai) žodžių ir žodžių formų srities dažninėje lentelėje dydį, t. y. maža vertė nurodo, kad tekstui būdinga didelė labai dažnų žodžių sritis, kitaip – tarp tekstą sudarančių žodžių ir jų formų labai dažnų žodžių ir jų formų yra daug, o didelė indikatoriaus a vertė reiškia, kad tekstui būdinga maža labai dažnų žodžių sritis arba kad tarp jų sudarančių žodžių ir žodžių formų labai dažnų žodžių ir jų formų esama nedaug. Indikatoriaus a vidurkis yra šiek didesnis nei mediana (žr. 20 lentelę), tad išskirčių galima tikėtis rasti didelių a verčių pusėj. Kitaip tariant, maža labai dažnų žodžių sritis MST tekstams būtų netipiškas atvejis.

20 lentelėje galima matyti, kad a verčių standartinis nuokrypis, lyginant su vidurkiu, yra nemažas. Tai reiškia, kad šio indikatoriaus vertės yra išsisklaidžiusios, kitaip sakant – MST būdinga a raiškos įvairovė – tekstai labai dažnų žodžių srities dydžio atžvilgiu ženkliai skiriasi tarpusavyje. Variacijos koeficiento vertė leidžia patvirtinti šias tendencijas, t. y. MST standartinis nuokrypis sudaro 19,7 % vidurkio reikšmės (20 lentelė), o tai rodo indikatoriaus verčių išsisklaidymą arba raiškos įvairovę, kuri, vėlgi, yra mažesnė nei AST (žr. 13 lentelę) ir PST (žr. 25 lentelę). Remiantis kvartilinio pločio verte (20 lentelė) galima sakyti, kad MST tekstai pagal labai dažnų žodžių ir žodžių formų srities dydžio parametą tarpusavyje skiriasi mažiaus nei PST (žr. 25 lentelę), bet daugiau nei AST (žr. 13 lentelę).

Pagal asimetrijos koeficiento reikšmę galima sakyti, kad MST dažnesnės didesnės a vertės arba, kitaip, mažesnė labai dažnų žodžių sritis (asimetrijos koeficiento vertė rodo didelę teigiamą asimetriją, žr. 20 lentelę). Ši tendencija MST tekstams yra stipresnė nei AST (žr. 13 lentelę) ir PST (žr. 25 lentelę). Galiausiai eksceso koeficientas rodo, kad MST būdingos a verčių išskirtys (žr. 20 lentelę), kitaip – netipiško dydžio labai dažnų žodžių sritys dažninėje lentelėje. Į išskirtis atsižvelgiama renkantis tolesnius tyrimo metodus.

4.2.1.3 R_1 indikatoriaus charakteristikos MST

Indikatoriaus R_1 , žyminčio žodžių ir žodžių formų įvairovę, vertės MST pasiskirsčiusios nuo 0,5625 iki 0,9250. Kitaip sakant, žodžių ir žodžių formų įvairovės raiška MST tekstams yra nuo 0,5625 (žymi mažiausią įvairovę) iki

0,9250 (žymi didžiausią įvairovę) (plg. AST – 0,2397 iki 0,9595 (plačiau žr. 4.1.1.3 sk.), o PST – nuo 0,5446 iki 0,9799 (plačiau žr. 4.3.1.3 sk.)). R_1 verčių vidurkis yra vos mažesnis nei mediana (žr. 20 lentelę), tad tai rodo, jog keletu išskirčių galima tikėtis mažų R_1 verčių pusėj. Kitaip sakant, itin maža žodžių ir žodžių formų įvairovė MST tekstams būtų netipiškas atvejis.

Kaip galima pastebėti 20 lentelėje, R_1 verčių standartinis nuokrypis, lyginant su vidurkiu, nėra labai nedidelis. Tai rodo, kad šio indikatorius vertės yra iš dalies artimos vidurkiui, kitaip – MST tekstai žodžių ir žodžių formų įvairovės atžvilgiu tarpusavyje skiriasi iš dalies. Variacijos koeficiento vertė patvirtina šią tendenciją, o lyginant tiriamuosius tekstinus galima tarti, kad R_1 verčių sklaida AST yra didesnė MST (AST tekstai tarpusavyje skiriasi labiau nei MST), o PST – mažesnė (PST tekstai tarpusavyje skiriasi mažiau nei MST) (žr. 13, 20 ir 25 lenteles). Kvartilinio pločio vertės (žr. 13, 20 ir 25 lenteles) atkartoja tą patį R_1 verčių sklaidos polinkį ir kartu tiriamųjų tekstinų tekstų tarpusavio panašumo-skirtumo polinkį.

Asimetrijos koeficiento reikšmė rodo, kad MST dažnesnės mažesnės R_1 vertės arba mažesnė žodžių ir žodžių formų įvairovė (asimetrijos koeficiento vertė rodo vidutinę neigiamą asimetriją, žr. 20 lentelę). Ši tendencija MST tekstams yra silpnesnė nei AST (žr. 13 lentelę) ir PST (žr. 25 lentelę), kitaip sakant, mažesnės žodžių ir žodžių formų įvairovės tendencijos AST ir PST yra stipresnės nei MST. Eksceso koeficiento vertė reiškia, kad MST tikėtina rasti šiek tiek išskirčių (netipiškų žodžių ir žodžių įvairovės atveju) (žr. 20 lentelę), į kurių buvimą svarbu atsižvelgti renkantis tolesnius analizės metodus.

4.2.1.4 MST tekstų RR_{mc} indikatorius charakteristikos

Indikatoriaus RR_{mc} , žyminčio žodžių ir žodžių formų koncentruotumą tekste, vertės MST pasiskirsčiusios nuo 0,8884 iki 0,9779. Tai rodo, kad žodžių ir žodžių formų koncentruotumas MST tekstuose varijuoja nuo 0,8884 (mažiausias koncentruotumas) iki 0,9779 (didžiausias koncentruotumas)) (plg. AST – 0,7398 iki 0,9858 (plačiau žr. 4.1.1.4 sk.), o PST – nuo 0,8321 iki 0,9932 (plačiau žr. 4.3.1.4 sk.)). Indikatoriaus vidurkis MST tekstams yra bemaž lygus medianai (20 lentelė), tad išskirčių reikėtų tikėtis nebent vos keletu. Kitaip sakant, didesnių ir mažesnių RR_{mc} verčių MST tekstams yra maždaug vienodai, neturėtų būti netipiškų raiškos atveju (nebent keletas).

20 lentelėje matoma, kad standartinis nuokrypis vos šiek tiek didesnis nei 0, o tai rodo, kad RR_{mc} vertės mažai skiriasi nuo vidurkio, t. y. žodžių ir žodžių formų koncentruotumo atžvilgiu MST tekstai tarpusavyje yra panašūs.

Variacijos koeficientas taip pat rodo RR_{mc} verčių artumą vidurkiui, t. y. tik labai mažus MST tekstų tarpusavio skirtumus pagal šį indikatorių. Lyginant su kitais tiriamaisiais tekstynais galima matyti, kad visų 3 jų tekstų vidinė diferenciacija žodžių ir žodžių formų koncentruotumo atžvilgiu yra itin maža (žr. 13, 20 ir 25 lenteles), nors AST tekstai pagal šį požymį tarpusavyje skiriasi truputį labiau nei MST, o PST – truputį mažiau nei MST. Kvartilinis plotis patvirtina šią tendenciją (žr. 13, 20 ir 25 lenteles).

Remiantis asimetrijos koeficiento verte galima sakyti, kad MST dažnesnės mažesnės RR_{mc} vertės arba, kitaip, mažesnis žodžių ir žodžių formų koncentruotumas (asimetrijos koeficiento vertė rodo didelę neigiamą asimetriją, žr. 20 lentelę). Šis MST tekstų polinkis yra silpnesnis nei AST (žr. 13 lentelę) ir PST (žr. 25 lentelę), kitaip sakant, mažesnis žodžių ir žodžių formų koncentruotumas AST ir PST yra tipiškėsnė nei MST. Eksceso koeficiento vertė parodė, kad MST tikėtina rasti šiek tiek išskirčių (netipiškų žodžių ir žodžių formų koncentruotumo atvejų) (žr. 20 lentelę). Į išskirčių buvimą kreipiamas dėmesys pasirenkant tolesnius tyrimo metodus.

4.2.1.5 MATTR indikatorius charakteristikos MST

MATTR, skirto informacijos srauto įvertinimui tekste, vertės MST tekstams pasiskirstė nuo 33,44 (žymi mažiausią informacijos srautą MST) iki 47,61 (žymi didžiausią informacijos srautą MST) (plg. AST – 14,76 iki 48,17 (plačiau žr. 4.1.1.5 sk.), o PST – nuo 26,94 iki 48,97 (plačiau žr. 4.3.1.5 sk.)). Indikatoriaus vidurkis MST tekstams mažesnis nei mediana (žr. 20 lentelę), tad tai reiškia, jog išskirčių turėtų būti mažų MATTR verčių pusė. Kitaip sakant, itin mažas informacijos srautas MST būtų išimtis, o ne tendencija.

Standartinis nuokrypis, kaip galima matyti 20 lentelėje, lyginant su vidurkiu yra nedidelis, tad MATTR vertės yra artimos vidurkiui. Kitaip sakant, pagal informacijos srauto raišką MST tekstai tarpusavyje mažai skiriasi. Variacijos koeficientas taip pat parodė, kad MATTR vertės yra artimos vidurkiui, kitaip – MST tekstai tarpusavyje gana panašūs. Kita vertus, pagal šį indikatorių MST tekstai tarpusavyje skiriasi labiau nei PST, bet mažiau nei AST (žr. 13, 20 ir 25 lenteles). Kvartilinis plotis patvirtina šią tendenciją (žr. 13, 20 ir 25 lenteles).

Pagal asimetrijos koeficiento vertę galima tarti, kad MST dažnesnės mažesnės MATTR vertės, kitaip – mažesnis informacijos srautas (asimetrijos koeficiento vertė rodo vidutinę neigiamą asimetriją, žr. 20 lentelę). Kita vertus, šis MST tekstų polinkis yra silpnesnis nei AST (žr. 13 lentelę) ir PST (žr. 25 lentelę), t. y. mažesnis informacijos srautas AST ir PST yra tipiškėsnė

nei MST. Galiausiai eksceso koeficiento vertė atskleidė, kad MST tikėtina rasti šiek tiek išskirčių (netipiškų informacijos srauto atvejų) (žr. 20 lentelę) ir į šią tendenciją svarbu atkreipti dėmesį renkantis tolesnius analizės metodus.

4.2.1.6 Teminės koncentracijos (TK) indikatorius charakteristikos MST

TK, vertinančio teminę koncentraciją tekste, vertės MST pasiskirstė nuo 0 (žymi mažiausią teminę koncentraciją MST, kurią tiksliau būtų įvardinti teminiu neutralumu) iki 0,92857 žymi didžiausią teminę koncentraciją MST) (plg. AST – 0,0505 iki 1,8833 (plačiau žr. 4.1.1.6 sk.), o PST – nuo 0 iki 2,4 (plačiau žr. 4.3.1.6 sk.)). TK vidurkis yra didesnis nei mediana (žr. 20 lentelę), o tai rodo, jog išskirčių turėtų būti didelių TK verčių pusė. Kitaip sakant, itin didelė teminė koncentracija MST tekstams nėra tipiška.

Standartinis nuokrypis, lyginant su vidurkiu, yra itin didelis, o tai reiškia, kad TK vertės MST yra itin išsisklaidžiusios vidurkio atžvilgiu. Kitaip sakant, pagal TK vertes MST tekstai tarpusavyje labai skirtingi. Variacijos koeficientas taip pat rodo didelę TK verčių sklaidą, žyminčią, kad MST tekstai pagal TK raišką yra labai skirtingi. Paminėtina, kad visiems 3 tiriamiesiems tekstynams būdinga didelė vidinė diferenciacija teminės koncentracijos raiškos atžvilgiu, nors AST tekstai pagal TK tarpusavyje panašesni nei MST, o PST tekstų tarpusavio skirtingumas TK atžvilgiu dar labiau išreikštas nei MST (žr. 13, 20 ir 25 lenteles). Pagal kvartilinio pločio vertę (apima 50 % verčių, atmetus 25 % didžiausių ir 25 % mažiausių) MST tekstai TK atžvilgiu tarpusavyje panašesni nei AST ir PST (žr. 13, 20 ir 25 lenteles).

Remiantis asimetrijos koeficiento verte galima sakyti, kad MST dažnesnės didesnės TK vertės, kitaip – didesnė teminė koncentracija (asimetrijos koeficiento vertė rodo didelę teigiamą asimetriją, žr. 20 lentelę). Kita vertus, šis MST tekstų polinkis yra truputį stipresnis nei PST (žr. 25 lentelę) ir gerokai stipresnis nei AST (žr. 13 lentelę). Kitaip sakant, didesnė teminė koncentracija MST yra šiek tiek būdingesnė nei PST ir gerokai tipiškesnė nei AST. Eksceso koeficiento vertė atskleidė, kad MST turėtų būti išskirčių (netipiškų teminės koncentracijos raiškos atvejų) (žr. 20 lentelę) ir į jas reikės atsižvelgti taikant tolesnius funkcinių stilių kalbinės raiškos analizės metodus.

4.2.1.7 MST tekstų indikatorių A ir D charakteristikos

Indikatoriai A ir D, skirti įvertinti, atitinkamai, teksto aktyvumą ir deskriptyvumą, yra vienas kitam priešinti, t. y. vieno jų raiškai stiprėjant, kito – silpnėja ir atvirksčiai, todėl jų statistinės savybės aptariamoms kartu. Indikatoriaus A vertės MST tekstams pasiskirsčiusios nuo 0,2785 (žymi mažiausią aktyvumo / dinamiškumo raišką) iki 0,8376 (žymi didžiausią aktyvumo / dinamiškumo raišką) (plg. AST – 0,0342 iki 1 (plačiau žr. 4.1.1.7 sk.), o PST – nuo 0,2727 iki 0,9619 (plačiau žr. 4.3.1.7 sk.)). Tuo tarpu indikatoriaus D vertės MST yra nuo 0,1624 (žymi mažiausią deskriptyvumo / aprašomumo raišką) iki 0,7215 (žymi didžiausią deskriptyvumo / aprašomumo raišką) (plg. AST – 0 iki 0,9658 (plačiau žr. 4.1.1.7 sk.), o PST – nuo 0,0381 iki 0,7273 (plačiau žr. 4.3.1.7 sk.)). A ir D vidurkiai MST tekstams truputį didesni nei jų medianos (žr. 20 lentelę), tad išskirčių turėtų būti didelių A ir D verčių pusėj. Kitaip sakant, itin didelė aktyvumo / dinamiškumo bei deskriptyvumo / aprašomumo raiška MST tekstams nėra tipiška.

Standartinis nuokrypis abiem indikatoriams yra vienodas ir parodo, kad A ir D vertėms būdingas verčių išsisklaidymas apie vidurkį (žr. 20 lentelę). Kitaip sakant, MST būdinga A ir D raiškos įvairovė. Variacijos koeficientas taip pat rodo indikatorių verčių išsisklaidymą, nors indikatoriui A jis yra mažesnis nei indikatoriui D (žr. 20 lentelę). Tai reiškia, kad pagal aktyvumo raišką MST tekstai tarpusavyje skiriasi labiau pagal deskriptyvumo. Kvartilinio pločio vertė nurodo, kad pagal aktyvumo ir deskriptyvumo raiškos įvairovę (A ir D verčių sklaidą) MST tekstai tarpusavyje skiriasi labiau nei PST, bet mažiau nei AST (žr. 13, 20 ir 25 lenteles).

Pagal asimetrijos koeficiento vertes galima sakyti, kad MST tekstams A ir D vertės yra pasiskirsčiusios gana tolygiai, t. y. maža asimetrija atskleidžia, kad didesnės ir mažesnės aktyvumo ir deskriptyvumo raiškos atveju šiame tekстыne yra daugmaž po lygiai (žr. 20 lentelę). Tuo tarpu AST ir PST būdingesnė mažesnė nei MST aktyvumo raiška bei didesnė – deskriptyvumo (žr. 13, 20 ir 25 lenteles). Pagal eksceso koeficiento vertę galima spręsti, kad MST neturėtų būti išskirčių (netipiškos raiškos atveju) arba būti vos keletas (žr. 20 lentelę).

4.2.1.8 Tarpveiksmažodinių atstumų (VA) indikatoriaus charakteristikos moksliniuose tekstuose

Indikatorius VA žymi teksto sintaksinės struktūros kompleksiskumą (Kubát, Matlach ir Čech, 2014, 42). Šio indikatoriaus vertės MST yra nuo 3,233 (priklauso paprasčiausios sintaksinės struktūros MST tekstui) iki 22,500 (priklauso kompleksiskiausios sintaksinės struktūros MST tekstui) (plg. AST – 3,098 iki 622,2 (plačiau žr. 4.1.1.8 sk.), o PST – nuo 2,652 iki 47,333 (plačiau žr. 4.3.1.8 sk.)). VA vidurkis MST tekstams yra didesnis nei mediana (žr. 20 lentelę), tad išskirčių (netipiškų VA raiškos atvejų), tikėtina, bus didelių VA verčių pusėje. Kitaip sakant, sintaksinės struktūros požiūriu itin sudėtingi tekstai turėtų būti greičiau išimtis nei tipiški atvejai.

Standartinis nuokrypis, lyginant su vidurkiu, nėra mažas, tad tai rodo, jog VA vertės MST yra išsisklaidžiusios vidurkio atžvilgiu (žr. 20 lentelę). Kitaip sakant, pagal sintaksinės struktūros kompleksiskumo požymį MST tekstai tarpusavyje yra skirtingi. Variacijos koeficientas taip pat rodo VA verčių sklaidą (VA verčių standartinis nuokrypis sudaro 19,7 % jo vidurkio reikšmės (20 lentelė)), rodančią, kad MST tekstai pagal VA raišką tarpusavyje yra skirtingi. Be to, ir visiems 3 tiriamiesiems tekstynams būdinga vidinė tekstų sintaksinės struktūros kompleksiskumo diferenciacija, nors MST, lyginant su kitais 2 tekstynais, ji yra mažiausia (plg. 13, 20 ir 25 lenteles). Pagal kvartilinio pločio vertę (apima 50 % indikatoriaus verčių, atmetus 25 % didžiausių ir 25 % mažiausių) MST tekstai sintaksinės struktūros kompleksiskumo požiūriu tarpusavyje panašesni nei AST ir kiek mažiau skiriasi nei (žr. 13, 20 ir 25 lenteles).

Pagal asimetrijos koeficiento vertę galima spręsti, kad MST dažnesnės didesnės VA vertės, t. y. sintaksinės struktūros požiūriu kompleksiškesni tekstai (asimetrijos koeficiento vertė rodo didelę teigiamą asimetriją, žr. 20 lentelę). Eksceso koeficiento vertė atskleidė, kad MST turėtų būti nemažai išskirčių, t. y. netipiškų teksto sintaksinės struktūros kompleksiskumo atvejų (žr. 20 lentelę), į kuriuos atsižvelgiama renkantis tolesnius tyrimo metodus.

4.2.2 Mokslinio stiliaus tekstų indikatorių koreliacija

Šiame poskyryje aptariama tiriamų indikatorių koreliacinė analizė, skirta statistinio ryšio tarp indikatorių poros įvertinimui. Tam pasirinktas Spirmeno (Spearman) koreliacijos koeficientas (Field, Miles ir Field, 2012, 223), nes MST kalbinę raišką žyminčių indikatorių vertės pasižymėjo išskirtimis

(Mukaka, 2012), t. y. pasitaikė itin didelių arba itin mažų reikšmių, išsiskiriančių iš bendro duomenų konteksto (Čekanavičius ir Murauskas, 2001, 36).

4.2.2.1 Išskirtys MST duomenyse

Kadangi išskirtys (netipiškos duomenų vertės – itin didelės arba itin mažos) yra svarbios tolesnių FS kalbinės raiškos požymių analizės metodų pasirinkimui, šiame skyriuje pristatoma išskirčių MST analizė. 21-oje lentelėje galima matyti, kad MST visų tirtų indikatorių vertėms būdingos išskirtys, nors jų ir mažiau nei AST (žr. 4.1.2.1 sk.) ir PST (žr. 4.3.2.1 sk.). Išskirčių intervalai indikatoriams buvo apskaičiuoti remiantis kvartilinio pločio (KP) statistika (plačiau žr. 3.1 sk.). Kitaip tariant, išskirtys – tai duomenų (indikatorių) vertės, dėl savo didumo ar mažumo išsiskiriančios iš bendro duomenų konteksto. Išskirčių skaičiavimų rezultatai pateikti 21-oje lentelėje.

21 lentelė. MST: indikatorių verčių išskirtys

Indikatorius	Išskirtys
VŽFI	Vertės, mažesnės nei 5,5 (rasta 4) ir didesnė nei 8,5 (rasta 1)
<i>a</i>	Vertės, didesnės nei 15 (rasta 16)
R_1	Vertės, mažesnės nei 0,67 (rasta 10)
RR_{mc}	Vertės, mažesnės nei 0,93 (rasta 49) ir didesnės nei 0,98 (rasta 3)
MATTR	Vertės, mažesnės nei 38 (rasta 14)
TK	Vertės, didesnės nei 0,5 (rasta 28)
A	Vertės, mažesnės nei 0,39 (rasta 5)
D	Vertės, didesnės nei 0,61 (rasta 4)
VA	Vertės, didesnės nei 7,4 (rasta 24)

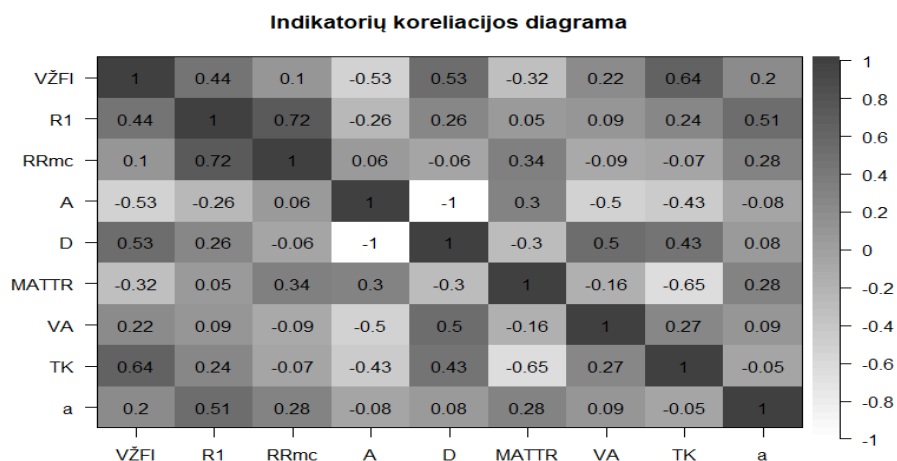
Kaip rodoma 21-oje lentelėje, indikatorius MST tekstams VŽFI vertės, mažesnės nei 5,5 (priklauso 4 MST tekstams), bei didesnės nei 8,5 (priklauso 1 MST tekstui), laikomos išskirtimis. Kitaip sakant, žodžiai ar žodžių formos, vidutiniškai trumpesni nei 5,5 simboliai ir vidutiniškai ilgesni nei 8,5 simboliai yra netipiški MST tekstams. Taip pat *a* vertės, didesnės nei 15 (rasta 16 MST tekstų, turinčių tokias vertes) priklauso išskirtims. Tai reiškia, kad labai dažnų žodžių sritys, kurių įvertis didesnis nei 15, yra netipiškos, dėl to, naudojant standartinius statistinius metodus, gali iškreipti analizės rezultatus. Pagal R_1 , žymintį (retesnių) žodžių ir žodžių formų įvairovę, netipiška MST atveju ši įvairovė yra, jei jos įvertis mažesnis nei 0,67. MST tekstuose

netipišką žodžių ir žodžių formų koncentruotumą rodo mažesnės nei 0,93 ir didesnės nei 0,98 RR_{mc} vertės.

Be to, MST netipiškas informacijos srautas (žymi MATTR indikatorius) yra toks, kurio įvertis yra mažesnis nei 38. Taip pat ir teminė koncentracija, jei ji yra didesnė nei 0,5, yra netipiška MST. Dar moksliniams tekstams netipiška yra itin mažas aktyvumas / dinamiškumas (A vertės, mažesnės nei 0,39) bei itin didelis deskriptyvumas / aprašomumas (D vertės, didesnės nei 0,61). Galiausiai ilgesni nei 7,4 vidutiniai atstumai tarp 2 tekste iš eilės einančių veiksmažodžių (žymi indikatorius VA) irgi nėra tipiški MST tekstams. Kadangi MST kontekste indikatorių vertėms būdingos išskirtys, ryšiams tarp indikatorių nustatyti ir įvertinti taikytas ne įprastas Pirsono, bet ranginis Spirmeno koreliacijos koeficientas (Mukaka, 2012; Lapinskas, 2003, 159).

4.2.2.2 Koreliacinės analizės rezultatai MST

MST indikatorių koreliacinės analizės rezultatai pateikti 8-ame paveiksle. Kaip jau buvo šiame darbe minėta, kuo didesnė koreliacijos koeficiento reikšmė (t. y. arčiau 1 arba -1) – tuo ryšys stipresnis. Kaip jau buvo minėta, kuo šviesesne spalva 8-ame paveiksle nuspalvintas kvadratinis, tuo stipresnis neigiamas ryšys sieja 2 indikatorius (vieno jų reikšmėms didėjant, kito – mažėja ir atvirkščiai). Lygiai taip pat, kuo tamsesne spalva nuspalvintas kvadratinis, tuo stipresnis teigiamas ryšys sieja 2 indikatorius (vieno jų reikšmėms didėjant/mažėjant, kito taip pat didėja/mažėja).



8 pav. Koreliacinės analizės diagrama MST

8-ame paveiksle pateiktoje diagramoje galima matyti, kad labai stipri koreliacija nustatyta 2 indikatorių poroms – RR_{mc} ir R_1 bei A ir D ¹⁰⁹. Vidutiniškai stiprus ryšys sieja 10 indikatorių porų: VŽFI ir R_1 ; VŽFI ir A; VŽFI ir D; R_1 ir a ; A ir VA; A ir TK; D ir VA; D ir TK; MATTR ir TK; VŽFI ir TK. Silpna koreliacija rasta tarp 12 indikatorių porų: VŽFI ir MATTR; VŽFI ir VA; VŽFI ir a ; R_1 ir A; R_1 ir D; R_1 ir TK; RR_{mc} ir MATTR; RR_{mc} ir a ; A ir MATTR; D ir MATTR; MATTR ir a ; VA ir TK. O labai silpna – tarp 8 indikatorių porų: RR_{mc} ir VŽFI; R_1 ir VA; D ir a ; VA ir a ; RR_{mc} ir VA; RR_{mc} ir TK; A ir a ; MATTR ir VA. 22-oje lentelėje pateikiami ryšių tarp indikatorių porų statistinio reikšmingumo įverčiai.

22 lentelė. MST: koreliacijos statistinis reikšmingumas (p -vertės)

	VŽFI	a	R_1	RR_{mc}	MATTR	TK	A	D	VA
VŽFI		0,00 0	0,00 0	0,000	0,000	0,00 0	0,00 0	0,00 0	0,00 0
a	0,000		0,00 0	0,000	0,000	0,11 10	0,01 61	0,01 61	0,00 28
R_1	0,000	0,00 0		0,000	0,1054	0,00 0	0,00 0	0,00 0	0,00 26
RR_{mc}	0,000	0,00 0	0,00 0		0,000	0,02 10	0,06 40	0,06 40	0,00 52
MATTR	0,000	0,00 0	0,10 54	0,000		0,00 0	0,00 0	0,00 0	0,00 0
TK	0,000	0,11 10	0,00 0	0,021 0	0,000		0,00 0	0,00 0	0,00 0
A	0,000	0,01 61	0,00 0	0,064 0	0,000	0,00 0		0,00 0	0,00 0
D	0,000	0,01 61	0,00 0	0,064 0	0,000	0,00 0	0,00 0		0,00 0
VA	0,000	0,00 28	0,00 26	0,005 2	0,000	0,00 0	0,00 0	0,00 0	

Kaip galima matyti 22 lentelėje, ryšiai tarp visų indikatorių porų buvo statistiškai reikšmingi ($p \leq 0,05$), išskyrus 4 indikatorių poras (a ir TK; R_1 ir MATTR; RR_{mc} ir A; RR_{mc} ir D). Šių 4 indikatorių porų ryšiai buvo statistiškai

¹⁰⁹ A ir D yra aiškiai priklausomi kintamieji, tad vieno jų būtų galima atsisakyti, tačiau, siekiant susidaryti bendrą vaizdą ir sekant čekų mokslininkų darbais (pavyzdžiui, Kubát (2015)), šiame tyrime etape pasiekiami abu indikatoriai. 5 sk. FS tarpusavio palyginimui taikant neparametrinius išvadų statistikos testus, iš šių dviejų indikatorių pasiekiamas tik indikatorius A.

neriškmingi ($p \geq 0,05$) (Field, Miles ir Field, 2012, 51—52)). Poskyriuose 4.2.2.2.1-4.2.2.2.3 pateikiami detalūs koreliacinės analizės MST rezultatai.

4.2.2.2.1 Labai stipri koreliacija tarp indikatorių

Duomenų analizė parodė, kad labai stipri statistiškai reikšminga koreliacija (Čekanavičius ir Murauskas, 2001, 126) randama tarp indikatorių RR_{mc} ir R_1 (0,72; žymi labai stiprią teigiamą koreliaciją) bei A ir D (-1; žymi labai stiprią neigiamą koreliaciją). Paminėtina, kad ryšys tarp RR_{mc} ir R_1 MST tekstams yra silpnesnis nei AST ir PST tekstams (žr. 5 ir 11 paveikslus). Indikatorius RR_{mc} leidžia įvertinti administracinio funkcinio stiliaus teksto žodžių ir žodžių formų koncentruotumą, o R_1 – (retesnių) žodžių ir žodžių formų įvairovę. Tarp RR_{mc} ir R_1 matome labai stiprų teigiamą ryšį, rodantį, kad didėjant/mažėjant žodžių ir žodžių formų koncentruotumui (kitais – tekste esant mažam skaičiui itin dažnai pasikartojančių žodžių, kai likusi dauguma pasikartoja gerokai rečiau), kartu didėja/mažėja ir (retesnių – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų įvairovė. Tai galima paaiškinti tuo, kad RR_{mc} įvertina visų teksto žodžių ir žodžių formų koncentraciją, o R_1 žymi tik retesnių (dažninėje lentelėje esančių žemiau h-taško) žodžių ir žodžių formų įvairovę. Indikatoriai A ir D žymi teksto „toną“ per aktyvumo (A) ir deskriptyvumo (D) dimensijas. Šie indikatoriai yra vienas kitam priešingi, tad natūralu, kad juos sieja itin stiprus neigiamas statistiškai reikšmingas ryšys, rodantis, kad tekste stiprėjant aktyvumo raiškai, deskriptyvumo raiška mažėja ir atvirkščiai.

4.2.2.2.2 Vidutiniškai stipri koreliacija tarp indikatorių

Vidutinio stiprumo statistiškai reikšminga koreliacija (Čekanavičius ir Murauskas, 2001, 126) nustatyta 10 indikatorių porų: $V\check{Z}FI$ ir R_1 ; $V\check{Z}FI$ ir A; $V\check{Z}FI$ ir D; R_1 ir a; A ir VA; A ir TK; D ir VA; D ir TK; MATTR ir TK; $V\check{Z}FI$ ir TK (žr. 10 pav.). 6 indikatorių poras sieja teigiamas ryšys ($V\check{Z}FI$ ir R_1 ; $V\check{Z}FI$ ir D; R_1 ir a; D ir VA; D ir TK; $V\check{Z}FI$ ir TK), t. y. vieno indikatoriaus vertėms didėjant/mažėjant, kito taip pat didėja/mažėja, o 4 – neigiamas ($V\check{Z}FI$ ir A; A ir VA; A ir TK; MATTR ir TK), t. y. vieno indikatoriaus vertėms didėjant, kito – mažėja ir atvirkščiai. Toliau plačiau apžvelgiami vidutinio stiprumo ryšiai, nustatyti tarp šių indikatorių porų.

Vidutinio teigiamas ryšys tarp $V\check{Z}FI$ ir R_1 rodo, kad didėjant/mažėjant teksto sudėtingumui/suprantamumui (arba kitaip – vidutiniam žodžio ir

žodžio formos ilgiui tekste) (VŽFI), MST tekste žodžių ir žodžių formų įvairovė (R_1) taip pat didėja/mažėja. Primintina, kad R_1 aprėpia tik retesnių, t. y. esančių žemiau h-taško, žodžių ir žodžių formų įvairovę (plačiau žr. 3.1.2.2 sk.). A ir VŽFI siejantis neigiamas ryšys reiškia, kad didėjant teksto aktyvumo (A) raiškai (kitaip – daugėjant tekste veiksmažodžių, lyginant su būdvardžiais), jo sudėtingumas/suprantamumas (VŽFI) mažėja (kitaip – vidutinis žodžių ir jų formų ilgis trumpėja) ir atvirkščiai. D ir VŽFI sieja teigiamas ryšys, kuris parodo, jog didėjant/mažėjant teksto deskriptyvumo (D) raiškai (t. y. daugėjant/mažėjant tekste būdvardžių, lyginant su veiksmažodžiais), jo sudėtingumas/suprantamumas (VŽFI) taip pat didėja/mažėja (t. y. ilgėja/trumpėja vidutinis žodžių ir žodžių formų ilgis).

Nustatytas teigiamas ryšys tarp indikatorių R_1 ir a rodo, kad kuo didesnė/mažesnė MST tekstų žodžių ir jų formų (retesnių) įvairovė, tuo juos sudarančių žodžių dažninėse lentelėse mažesnė / didesnė labai dažnų žodžių ir žodžių formų sritis. Neigiamas ryšys tarp VA ir A nurodo, kad kuo MST tekstų sintaksinė struktūra sudėtingesnė (kitaip tariant, kuo šiuose tekstuose atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis), tuo mažesnė juose aktyvumo raiška, t. y. tuo mažiau juose, lyginant su būdvardžiais, yra veiksmažodžių bei atvirkščiai. Indikatorius A ir TK siejantis neigiamas ryšys reiškia, kad didėjant MST teksto aktyvumo raiškai, jo teminė koncentracija mažėja ir atvirkščiai. Kitaip sakant, kuo daugiau tekste veiksmažodžių, lyginant su būdvardžiais, tuo jame mažiau retesnių, pagrindinėms kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu bemaž prilygstančių funkciniam. O vidutinio stiprumo teigiamas ryšys tarp VA ir D pažymi, kad kuo teksto sintaksinė struktūra sudėtingesnė/paprastesnė (kitaip tariant, kuo tekste atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis/mažesnis), tuo jame deskriptyvumo raiška didesnė/mažesnė (t. y. tuo daugiau/mažiau tekste būdvardžių, lyginant su veiksmažodžiais).

D ir TK siejantis vidutinio stiprumo teigiamas ryšys parodo, jog didėjant/mažėjant deskriptyvumo raiškai MST tekste, šio teksto teminė koncentracija taip pat didėja/mažėja. Kitaip sakant, kuo tekste daugiau būdvardžių, lyginant su veiksmažodžiais, tuo jame daugiau funkciniam žodžiams dažnumu nedaug nusileidžiančių retesnių, pagrindinėms kalbos dalims priklausančių, žodžių ir žodžių formų. Neigiamas ryšys tarp MATTR ir TK reiškia, kad kuo didesnis informacijos srautas (t. y. kuo didesnis skaičius skirtingų žodžių ir žodžių formų) būdingas tekstui, tuo mažesnė jo teminė koncentracija (kitaip – tuo mažiau jame iš esmės retesnių turinio/teminių, tačiau šiame tekste savo dažnumu funkciniam bemaž prilygstančių, žodžių ir žodžių formų) ir atvirkščiai. Galiausiai vidutinio stiprumo teigiamas ryšys tarp TK ir VŽFI rodo, kad didėjant/mažėjant teminei koncentracijai (TK), teksto

sudėtingumas / suprantamumas (VŽFI) taip pat didėja/mažėja, t. y. ilgėja/trumpėja vidutinis žodžių ir jų formų ilgis.

4.2.2.2.3 Silpna ir labai silpna koreliacija tarp indikatorių

Silpna koreliacija nustatyta 12 indikatorių porų: VŽFI ir MATTR; VŽFI ir VA; VŽFI ir *a*; R₁ ir A; R₁ ir D; R₁ ir TK; RR_{mc} ir MATTR; RR_{mc} ir *a*; A ir MATTR; D ir MATTR; MATTR ir *a*; VA ir TK. Iš jų teigiamas ryšys (vieno indikatoriaus vertėms didėjant/mažėjant, kito irgi didėja/mažėja) rastas 9 indikatorių poroms (VŽFI ir VA; VŽFI ir *a*; R₁ ir D; R₁ ir TK; RR_{mc} ir MATTR; RR_{mc} ir *a*; A ir MATTR; MATTR ir *a*; VA ir TK), neigiamas – 3 poroms (VŽFI ir MATTR; R₁ ir A; D ir MATTR) Labai silpna koreliacija nustatyta 8 indikatorių poras, iš jų 4 poras sieja teigiamas ryšys (RR_{mc} ir VŽFI; R₁ ir VA; D ir *a*; VA ir *a*), o kitas 4 – neigiamas (RR_{mc} ir VA; RR_{mc} ir TK; A ir *a*; MATTR ir VA).

Koreliacija tarp indikatorių MST kontekste buvo statistiškai nereikšminga tarp TK ir *a*; R₁ MATTR; RR_{mc} ir A; RR_{mc} ir D. Šių porų ryšio statistinį reikšmingumą žyminčios *p* vertės yra mažesnės nei 0,05 (22 lentelė). Tad galima sakyti, kad MST tekstų atžvilgiu disertacijoje naudojami kalbinę raišką žymintys indikatoriai, išskyrus minėtas 4 poras, tarp kurių statistiškai reikšmingo ryšio neužfiksuota, yra stipriau ar silpniau susiję, t. y. turi įtakos vienas kito verčių kitimui.

4.2.3 Pagrindinių komponentių analizės (PKA) rezultatai MST

4.2.3.1 Indikatorių svarba reprezentuojant mokslinio stiliaus kalbinę raišką

Siekiant išskirti indikatorius bei jų kombinacijas, geriausiai apibūdinančius MST kalbinę raišką, buvo atlikta pagrindinių komponentių analizė (PKA; plačiau žr. 3.3 sk.). 23-oje lentelėje pateikiama MST kalbinę raišką reprezentuojančių indikatorių PKA rezultatų santrauka (lentelėje „PK1“, ..., „PK9“ žymi pagrindines komponentes). Pirmoji pagrindinė komponentė apibūdina didžiausią dalį duomenų verčių variacijos (žr. 23-ą lentelę, eilutę „Variacijos proporcija, %“). Kuo didesnę duomenų variantiškumo dalį ši komponentė apibūdina, tuo daugiau informacijos apie tiriamą objektą (MST kalbinės raiškos požymius) ji išreiškia. Antroji komponentė reprezentuoja didžiausią dalį likusios indikatorių verčių dispersijos, t. y. iš to, kas liko nuo

pirmosios pagrindinės komponentės. Pagrindinių komponentių yra tiek, kiek kintamųjų (9, pagal tiriamų indikatorių skaičių; žr. 23-ą lentelę).

Kaip rodo 23-a lentelė, pirmos trys pagrindinės komponentės paaiškina didžiąją dalį visos duomenų dispersijos. Pirmoji pagrindinė komponentė PK1 išreiškia 36,43 proc. indikatorių verčių variacijos, antroji pagrindinė komponentė PK2 – 24,72 proc., o trečioji – 12,08 proc. (žr. 23-os lentelės eilutę „Variacijos proporcija, %“). Kartu jos apibūdina 73,23 proc. visos indikatorių verčių variacijos ($PK1 + PK2 + PK3 = 36,43 + 24,72 + 12,08 = 73,23$; žr. 23-os lentelės eilutę „Sukauptinė variacijos proporcija, %“). Likusios 6 komponentės kiekviena atskirai išreiškia vis mažesnę dalį duomenų variacijos (žr. 23-os lentelės eilutę „Variacijos proporcija, %“). Kartu jos apibūdina 26,77 proc. indikatorių reikšmių ($PK4 + PK5 + PK6 + PK7 + PK8 + PK9 = 8,47 + 8,01 + 5,16 + 3,68 + 1,44 + 0,00 = 26,77$; žr. 23-os lentelės eilutę „Variacijos proporcija, %“). Visų kalbinę raišką žyminčių indikatorių reikšmių atvaizdavimui pakanka 7 komponentių (sudėjus pirmas 7 komponentes, jos visos kartu išreiškia 98,56 proc. visos reikšmių variacijos). Tad tolesnei analizei pasiliekomos tik didžiausią indikatorių reikšmių variacijos dalį paaiškinančios komponentės PK1, PK2 ir PK3. Jų tikrinių vektorių reikšmės (rodo, kokią dalį pradinių duomenų variacijos paaiškina komponentės (Janilionis, Morkevičius, ir Rauleckas 2008, 171); žr. 23 lentelę) yra didesnė nei 1, tad tai rodo, jog šios komponentės reprezentuoja didesnę dalį indikatorių reikšmių variacijos nei vienas kuris indikatorius iš pirminio (iki analizės) indikatorių verčių rinkinio (Kassambara, 2017, 18).

23 lentelė. PKA rezultatų santrauka MST: duomenų variacijos proporcijos, kurias paaiškina komponentės

	PK1	PK2	PK3	PK4	PK5	PK6	PK7	PK8	PK9
Tikrinių vektorių reikšmės	3,28	2,23	1,09	0,76	0,72	0,46	0,33	0,13	0,000
Variacijos proporcija (%)	36,43	24,72	12,08	8,47	8,01	5,16	3,68	1,44	0,000
Sukauptinė variacijos proporcija (%)	36,43	61,15	73,23	81,71	89,72	94,88	98,56	100,000	100,000

Remiantis PKA rezultatais (23 lentelė), toliau buvo analizuojama, kurie indikatoriai tiriant mokslinio funkcinio stiliaus kalbinę raišką išreiškia

didžiausią dalį duomenų reikšmių, t. y. yra svarbiausi arba, kitaip sakant, geriausiai apibūdina šio stiliaus raišką. PKA rezultatai pateikiami dvimatėje diagramoje programos viduje, kur duomenų taškas (t. y. kiekviena kiekvieno indikatorius vertė) turi dvi koordinates (horizontalioje ašyje ir vertikalioje ašyje). Indikatorius svarba pagrindinėje komponentėje įvertinama pagal kampo tarp šių koordinatų didumą (Husson, Josse, Le ir Mazet, 2018) (plačiau žr. 4.1.3.1 sk.). Kuo didesnė ši vertė, tuo indikatorius konkrečioje komponentėje yra svarbesnis, kitaip – tuo jis komponentėje geriau reprezentuojamas. Kiekvieno indikatorius indėlis atskirose komponentėse skaičiuojamas procentais įvertinant indikatorius svarbą visoje komponentėje (Kassambara, 2017, 20) (plačiau žr. 4.1.3.1 sk.). Kuo ši procentinė išraiška didesnė, tuo indikatorius indėlis konkrečioje komponentėje yra didesnis.

24 lentelė. PKA rezultatų santrauka MST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis

	PK1		PK2		PK3	
	Indėlis (%)	Svarba (\cos^2)	Indėlis (%)	Svarba (\cos^2)	Indėlis(%)	Svarba (\cos^2)
VŽFI	18,049	0,592	1,207	0,027	9,512	0,103
<i>a</i>	0,678	0,022	19,511	0,434	1,000	0,011
R_1	6,098	0,200	27,682	0,616	4,871	0,053
RR_{mc}	0,007	0,000	31,519	0,701	1,706	0,019
MATTR	5,947	0,195	16,792	0,374	15,094	0,164
TK	13,521	0,443	2,622	0,058	21,511	0,234
A	24,137	0,791	0,092	0,002	9,154	0,100
D	24,137	0,791	0,092	0,002	9,154	0,100
VA	7,426	0,243	0,483	0,011	27,997	0,304

Kaip galima matyti 24-oje lentelėje, PK1 pagal \cos^2 verčių didumą geriausiai reprezentuojami yra indikatoriai A ir D (primintina, kad po duomenų transformavimo arba „pasukimo“ komponentę sudaro ne vieno indikatorius vertės, o indikatorių verčių kombinacija). A ir D įvertina tekstą per aktyvumo ir deskriptyvumo dimensijas, kitaip – teksto „toną“. Šių indikatorių indėlis šioje komponentėje taip pat yra didžiausias (žr. 24-os lentelę). A ir D kartu apibūdina 48,27 proc. PK1 duomenų variacijos ($A(24,137) + D(24,137) = 48,27$ proc.; žr. 24 lentelę). PK1 išreiškia 36,43 proc. visų indikatorių verčių sklaidos. Tad PK1 kombinacija A ir D nusako 17,6 proc. visų indikatorių verčių variacijos. Kitų indikatorių svarba ir indėlis

šioje komponentėje yra mažesni, o mažiausiai svarbus bei kartu mažiausiai prisidedantis apibūdinant mokslinio stiliaus tekstus) yra indikatorius RR_{mc} , vertinantis žodžių ir žodžių formų koncentraciją tekste. Taigi aktyvumo (A) ir deskriptyvumo (D) požymiai iš visų indikatorių reprezentuoja daugiausiai PK1 esančių verčių, o žodžių ir žodžių formų koncentracijos (RR_{mc}) požymis šioje komponentėje yra mažiausiai svarbus, nes nusako itin mažą dalį visų PK1 esančių indikatorių verčių.

Komponentė PK2 reprezentuoja 24,72 proc. visų indikatorių reikšmių (žr. 23 lentelę). Pagal svarbą ir indėlį šioje komponentėje išsiskiria indikatoriai RR_{mc} ir R_1 , turintys didžiausias \cos^2 vertes (žr. 24 lentelę). RR_{mc} yra žodžių ir jų formų koncentracijos indikatorius, o R_1 – žodžių ir žodžių formų įvairovės rodiklis. Abu kartu šie indikatoriai apibūdina 59,2 proc. PK2 indikatorių verčių variacijos ($RR_{mc}(31,519) + R_1(27,682) = 59,2$ proc.; žr. 24 lentelę). Kadangi PK2 reprezentuoja 24,72 proc. visų indikatorių reikšmių, tai RR_{mc} ir R_1 kombinacija išreiškia 14,63 proc. visos indikatorių reikšmių sklaidos. Kitų indikatorių svarba ir indėlis mažesni. Mažiausiai apibūdinant MST komponentėje PK2 prisideda indikatoriai A ir D, turintys mažiausias \cos^2 reikšmes. Tad žodžių ir žodžių koncentracija bei įvairovė (RR_{mc} ir R_1) yra svarbiausi požymiai PK2 kontekste apibūdinant MST, o aktyvumas ir deskriptyvumas (A ir D) – mažiausiai svarbūs.

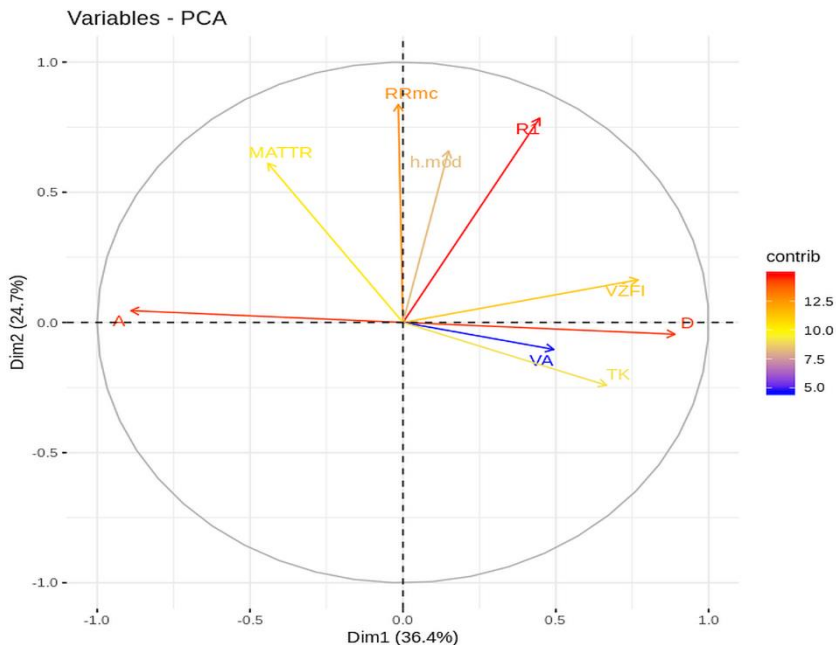
Paskutinėje analizuotoje pagrindinėje komponentėje PK3 išsiskiria indikatoriai – VA ir TK. Pirmasis nusako teksto sintaksinės struktūros sudėtingumą, o antrasis – įvertina teksto teminę koncentraciją. Abu kartu VA ir TK indikatoriai nusako 49,5 proc. KP3 indikatorių verčių variacijos ($VA(27,997) + TK(21,511) = 49,5$ proc.). PK3 reprezentuoja 12,08 proc. visos indikatorių verčių sklaidos, tad VA ir TK nusako 5,98 proc. visos indikatorių verčių variacijos. Kitų indikatorių svarba ir indėlis mažesni. Mažiausias indėlis ir svarba apibūdinant publicistinį stilių komponentėje PK3 priklauso indikatoriui a, kurio \cos^2 reikšmė lygi 0,011. Tad teksto sintaksinės struktūros kompleksiskumas (VA) ir teminė koncentracija yra svarbiausi požymiai PK3 kontekste apibūdinant MST, o indikatorius a, apibūdinantis labai dažnų žodžių ir jų formų srities dažninėje lentelėje dydį, – mažiausiai svarbus.

4.2.3.2 Indikatorių indėlis reprezentuojant mokslinio stiliaus kalbinę raišką

Ištyrus atskirų indikatorių svarbą charakterizuojant MST kalbinę raišką, atlikta ir indikatorių indėlio apibūdinant MST kalbinę raišką analizė, kurios rezultatai pateikti 9-ame ir 10-ame paveiksluose. Kuo indikatorius reprezentuojančių rodyklių spalva raudonesnė, tuo jų indėlis apibūdinant MST

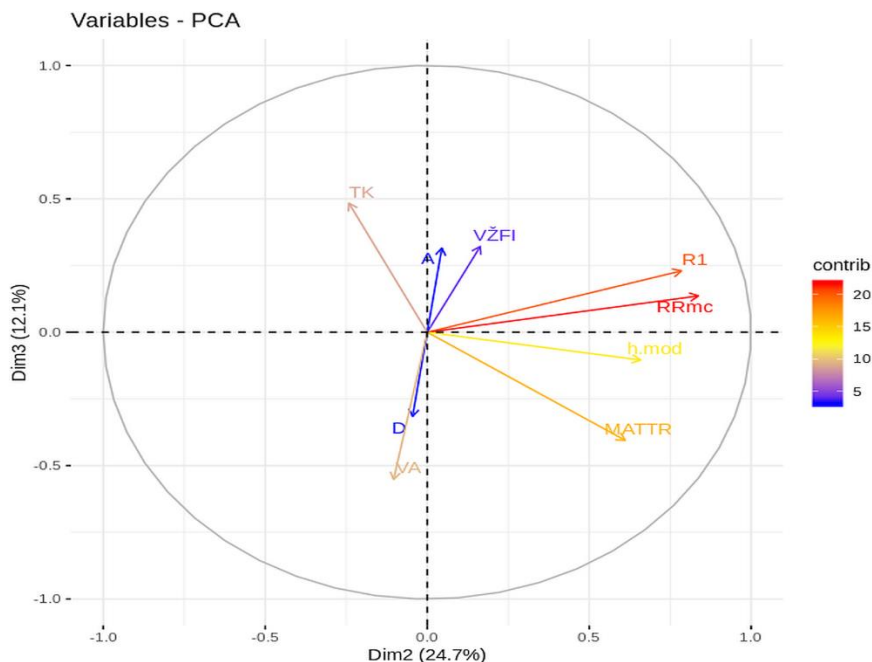
kalbinę raišką yra didesnis. Ir kuo indikatorius reprezentuojančių rodyklių spalva mėlynesnė/labiau violetinė, tuo jų indėlis apibūdinant MST kalbinę raišką yra mažesnis. Be to, kuo šių rodyklių smaigaliai yra arčiau paveikslė pavaizduoto apskritimo perimetro, tuo indikatorių indėlis taip pat didesnis, o kuo smaigaliai yra arčiau apskritimo centro, tuo indikatorių indėlis yra mažesnis.

9-ame paveikslė pavaizduoti atskirų indikatorių indėliai apibūdinant MST kalbinę raišką bendrame pirmųjų dviejų pagrindinių komponentų PK1 ir PK2 kontekste. Tad galima matyti, kad didžiausias indėlis apibūdinant MST PK1 ir PK2 kontekste priklauso indikatoriams A, D, RR_{mc} ir R_1 (9 pav.) – juos vaizduojančių rodyklių spalva yra raudoniausia, o tų pačių rodyklių smaigaliai baigiasi arčiausiai paveikslė pavaizduoto apskritimo perimetro. Kitaip sakant, PK1 ir PK2, reprezentuojančiose daugiau nei pusę (61,15 proc.) MST kalbinę raišką apibūdinančių indikatorių verčių variacijos (23 lentelė), teksto „tonas“ (aktyvumo-deskriptyvumo dimensijos, t. y. indikatoriai A ir D), žodžių ir jų formų koncentracija (RR_{mc}) bei žodžių ir jų formų įvairovė (R_1) labiausiai prisideda charakterizuojant šį funkcinį stilių. Kitų indikatorių (VŽFI, MATTR, TK) (rodyklių, reprezentuojančių šiuos indikatorius, spalva yra labiau oranžinė nei raudona ar net geltona, o smaigalys baigiasi toliau nuo paveikslė pavaizduoto apskritimo perimetro nei A, D, RR_{mc} ir R_1). Mažiausias indėlis apibūdinant MST priklauso VA. Tai reiškia, kad teksto sintaksinė struktūra ir jos kompleksškumą žymintis VA nėra charakteringiausias MST požymis.



9 pav. MST: indikatorių indėlis (PK1 ir PK2)

Indikatorių svarba pagal jų indėlį pagrindinėje komponentėje PK3 pavaizduota 10-ame paveiksle. Kaip ir indikatorių svarbos analizės atveju, taip ir tiriant indikatorių indėlį reprezentuojant MST, dėl kintamųjų reikšmių atvaizdavimo tikslumo antrąja ašimi dvimatei diagramai pasirenkama antroji pagrindinė komponentė PK2, jau aptarta aukščiau. Tad 10-ame paveiksle X (horizontalioje) ašyje pavaizduota PK2 (ji reprezentuoja didesnę procentą visų kintamųjų reikšmių), o Y (vertikalioje) ašyje pateikiama PK3.



10 pav. MST: indikatorių indėlis (PK2 ir PK3)

Taigi kaip matyti 10-ame paveiksle, bendrame komponentų PK2 ir PK3 kontekste indikatoriai RR_{mc} , R_1 , VA ir TK daugiausiai prisideda apibūdinant MST – juos reprezentuojančios rodyklės yra raudoniausios, o jų smaigaliai baigiasi arčiausiai paveiksle pavaizduoto apskritimo perimetro. Indikatoriai RR_{mc} , R_1 apibūdina tekstą per žodžių ir žodžių formų įvairovės-koncentracijos dimensijas. Jų indėlis mokslinio funkcinio stiliaus tekstų apibūdinimui aptartas anksčiau, analizuojant 9-ą paveikslą.

Indikatorius VA įvertina teksto sintaksinės struktūros kompleksumą per vidutinį atstumą tarp dviejų iš eilės einančių veiksmažodžių, o TK apibūdina, kaip tekstas yra sukonzentruotas į savo temą pagal dažnumu funkciniam žodžiams ir jų formoms beveik prilygstančių turinio žodžių ir jų formų skaičių. Analizuojant duomenis galima pastebėti, kad kitų indikatorių indėlis mažesnis (rodyklių smaigaliai arčiau apskritimo centro, o jų spalvos – tolimesnės raudonos spalvos spektrui). Mažiausiai bendrame PK2 ir PK3 kontekste apibūdinant MST prisideda indikatoriai A ir D (aktyvumo-deskriptyvumo dimensijos). Kitaip sakant, aktyvumas ir deskriptyvumas nėra charakteringiausi MST požymiai, nes reprezentuoja tik labai mažą dalį į PK2 ir PK3 patekusių indikatorių reikšmių.

4.3 Publicistinio stiliaus tekstų kalbinės raiškos požymių kiekybinė analizė

4.3.1 Aprašomoji (pirminė) publicistinių tekstų požymių analizė

Siekiant apibendrinti PST kalbinės raiškos požymių statistines savybes, atlikta aprašomoji duomenų analizė (Casas, 2016). Bazinių statistinių charakteristikų santrauka kiekvienam tiriamam indikatoriumi (indikatorių vertės buvo suskaičiuotos kiekvienam publicistinio stiliaus tekstyno tekstui) pateikiama 25-oje lentelėje. Detalūs rezultatai pateikiami 4.3.1.1–4.3.1.8 sk.

25 lentelė. PST: bazinių statistinių charakteristikų santrauka

Indikatorius	Vidurkis	Mediana	Standartinis nuokrypis	Variacijos koef.	Kvartilinis plotis	Asimetrijos koef.	Eksceso koef.
VŽFI	6,374	6,382	0,3836	0,0602	0,516	-0,051	3,0
<i>a</i>	14,36	13,92	3,5543	0,2476	4,332	0,918	5,1
R ₁	0,9159	0,9183	0,0232	0,0253	0,028	-1,630	13,6
RR _{mc}	0,9743	0,9751	0,0077	0,0079	0,009	-2,289	24,0
MATTR	46,04	46,22	1,2704	0,0276	1,293	-2,561	19,3
TK	0,2586	0,1930	0,2470	0,9550	0,310	1,504	6,3
A	0,7328	0,7384	0,0768	0,1048	0,101	-0,539	3,7
D	0,2672	0,2616	0,0768	0,2874	0,101	-0,539	3,7
VA	4,661	4,489	1,1784	0,2528	1,033	9,433	252,3

Vidurkis ir mediana žymi tipiškas kiekvieno indikatoriaus vertes, standartinis nuokrypis, variacijos koeficientas ir kvartilinis plotis leidžia įvertinti indikatorių verčių sklaidą, t. y. kaip labai publicistiniai tekstai pagal konkretų indikatorių skiriasi tarpusavyje. Pagal asimetrijos koeficientą galima įvertinti, kokios konkrečiau indikatoriaus vertės yra dažnesnės – didesnės ar mažesnės. Eksceso koeficientas yra skirtas išskirčių duomenyse įvertinimui bei tam tikrų netipiškų atvejų pagal konkrečiau indikatoriaus raišką masto nustatymui. Tai turi įtakos tolesnių tyrimo metodų pasirinkimui.

4.3.1.1 Vidutinio žodžių ir žodžių formų ilgio (VŽFI) indikatoriaus charakteristikos PST tekstuose

VŽFI, paprastai naudojamo tekstų sudėtingumui (angl. *readability*) įvertinti (van Oosten, Tanghe ir Hoste, 2010), statistinių savybių, gautų atlikus aprašomąją duomenų analizę, santrauka pateikta 25-oje lentelėje. PST tekstams VŽFI reikšmės pasiskirsčiusios intervale nuo 5,013 (mažiausia reikšmė, priklausanti tekstui, nepriskirtam jokiai teminei kategorijai (*Nenurodyta kategorija*)) iki 7,945 (didžiausia reikšmė, priklausanti skilties *Mokslas* tekstui). Kitaip tariant, vidutinis žodžių formos ilgis konkrečioms šio tekstyno tekstams yra nuo 5,013 iki 7,945 simbolių (plg.: AST – nuo 5,373 iki 9,209 simbolių (plačiau žr. 4.1.1.1 sk.), o MST – nuo 5,208 iki 11,536 simbolių (plačiau žr. 4.2.1.1 sk.)). Indikatoriaus vidurkio ir medianos vertės yra vienodos – 6,323. Tai rodo, kad indikatoriaus verčių dažniai yra pasiskirstę simetriškai, kitais žodžiais tariant, tekstų, kuriems būdingi vidutiniškai ilgesni žodžiai ar jų formos, yra tiek pat, kiek ir tų, kuriems būdingi vidutiniškai trumpesni žodžiai ar jų formos. Tai rodo, kad PST tekstams netipiško vidutinio ilgio žodžiai nėra būdingi.

Kaip matyti iš 25 lentelės, VŽFI verčių standartinio nuokrypio vertė, lyginant su vidurkiu yra vidutiniškai didelė, o tai reiškia, kad indikatoriaus reikšmės yra vidutiniškai nutolusios nuo vidurkio. Kitaip sakant, pagal vidutinį žodžio ar žodžio formos ilgį PST tekstai iš dalies skiriasi tarpusavyje. Variacijos koeficiento reikšmė (0,0602; žr. 25 lentelę) taip pat nurodo šią tendenciją. Kvartilinio pločio vertė rodo, kad VŽFI verčių sklaida PST yra didesnė nei AST, bet mažesnė nei MST (žr. 13, 20 ir 25 lenteles). Kitais žodžiais tariant, pagal vidutinio žodžio ar žodžio formos ilgio parametą PST tekstai tarpusavyje skiriasi labiau nei AST, bet mažiau nei MST.

Asimetrijos koeficiento vertė VŽFI indikatoriui yra silpnai neigiama (-0,051). Tai rodo, kad indikatoriaus vertės yra pasiskirsčiusios beveik simetriškai (Bulmer, 1979, 63), t. y. tarp indikatoriaus verčių mažesnės nei vidurkis vertės yra tik šiek tiek dažnesnės. Tai reiškia, kad PST tekstams šiek tiek būdingesni, nors tas polinkis ir nedidelis, vidutiniškai trumpesni nei vidurkis žodžiai ir žodžių formos. Eksceso koeficiento vertė (3; žr. 25 lentelę) rodo, kad duomenų histograma yra „varpo“ formos (reikšmės simetriškai pasiskirsčiusios apie vidurkį), juose nėra išskirčių, t. y. itin didelių arba itin mažų reikšmių, išsiskiriančių iš bendro duomenų konteksto (Čekanavičius ir Murauskas, 2001, 36). Tai rodo, kad netipiški vidutiniai žodžių ir žodžių formų ilgiai PST tekstams nėra būdingi.

4.3.1.2 Indikatoriaus *a* charakteristikos PST

Indikatorius *a* žymi labai dažnų žodžių ir žodžių formų srities (t. y. iki h-taško, plačiau žr. 1.4.2.1 sk.) dydį dažninėje teksto sudarančių žodžių lentelėje. Atliktus aprašomąją duomenų analizę, įvertintos šio indikatoriaus bazinės statistinės savybės PST tekstuose. Aprašomosios duomenų analizės santrauka pateikta 25-ioje lentelėje. Indikatoriaus *a* reikšmės PST tekstams pasiskirstė nuo 3,59 (mažiausia reikšmė; priklauso *Sporto* skilties tekstui) iki 43,22 (didžiausia reikšmė; priklauso *Gyvenimo* skilties tekstui) (plg.: AST – 1,576 iki 39,50 (plačiau žr. 4.1.1.2 sk.), o MST – 5,450 iki 25,429 (plačiau žr. 4.2.1.2 sk.)). Mažiausia *a* reikšmė rodo, kad šią reikšmę įgijusiam tekstui būdinga didelė labai dažnų žodžių ir žodžių formų sritis – gausu labai dažnų žodžių ir žodžių formų. Didžiausia *a* reikšmė rodo, kad ją įgijusiam tekstui būdinga maža labai dažnų žodžių sritis, t. y. tekstui nebūdinga labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų gausa.

Indikatoriaus reikšmių PST tekstams vidurkis didesnis nei mediana (žr. 25 lentelę). Tai rodo, kad išskirčių, t. y. itin didelių arba mažų verčių, išsiskiriančių iš visų indikatoriaus verčių, jas išrikiavus pagal dydį (Čekanavičius ir Murauskas, 2001, 36), galima tikėtis didelių verčių pusėj. Kitaip tariant, itin maža labai dažnų žodžių ir jų formų sritis nėra tipiška PST tekstams. Indikatoriaus reikšmių standartinio nuokrypio vertė yra daug didesnė už 0 (3,5543), o tai reiškia, kad *a* reikšmės yra labai išsisklaidžiusios. Kitaip tariant, PST tekstai pagal labai dažnų žodžių ir žodžių formų parametą tarpusavyje yra skirtingi. Variacijos koeficiento reikšmė (0,2476) patvirtina šią tendenciją. Tad *a* reikšmės nėra artimos vidurkiui, t. y. PST tekstai pagal *a* reikšmių mastą gerokai skiriasi tarpusavyje. Kitaip tariant, šio stiliaus tekstams būdinga labai dažnų ir retesnių žodžių ir jų formų proporcijų įvairovė. Pagal kvartilinį plotį *a* vertės PST yra labiau išsisklaidžiusios nei AST ir MST (žr. 13, 20, ir 25 lenteles). Tai rodo, kad pagal labai dažnų žodžių ir žodžių formų srities dydžio parametą PST tekstai tarpusavyje skiriasi labiau nei AST ir MST.

Asimetrijos koeficiento vertė *a* indikatoriui yra teigiama (0,918; žr. 25 lentelę). Kitaip sakant, PST dažnesnės didesnės *a* vertės, žyminčios mažesnę labai dažnų žodžių sritį. Eksceso koeficiento vertė yra 5,1 ir tai rodo, kad duomenų sklaida apie vidurkį yra didesnė nei normaliosios kreivės. Kitaip tariant, tarp *a* verčių esama netipiškai didelių arba mažų ir tai reikės atsižvelgti renkantis tolesnės analizės metodus.

4.3.1.3 R_1 indikatoriaus charakteristikos PST

R_1 indikatorius (Kubát, Matlach, ir Čech, 2014, 14) šioje disertacijoje – žodžių ir jų formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovės rodiklis. Siekiant pažinti indikatoriaus bazines statistines savybes, buvo atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 25-oje lentelėje. R_1 reikšmės pasiskirsčiusios nuo 0,5446 (mažiausia reikšmė, rodanti mažiausią žodžių ir jų formų įvairovę; priklauso tekstui, nepriskirtam jokiai teminei kategorijai (*Nenurodyta kategorija*)) iki 0,9799 (didžiausia reikšmė, rodanti didžiausią žodžių ir jų formų įvairovę; tekstas, įgijęs šią reikšmę, priklauso *Sporto* skilčiai) (plg.: AST – 0,2397 iki 0,9595 (plačiau žr. 4.1.1.3 sk.), o MST – nuo 0,5625 iki 0,9250 (plačiau žr. 4.2.1.3 sk.)). Indikatoriaus verčių PST tekstams vidurkis (0,9159) yra mažesnis nei mediana (0,9183). Tai rodo, kad išskirtys, t. y. itin didelės arba mažos indikatoriaus vertės yra tikėtinos didelių verčių pusėj. Kitaip tariant, itin didelė (retesnių) žodžių ir žodžių formų įvairovė PST, tikėtina, bus tik pavieniai atvejai, o ne tendencija.

Indikatoriaus verčių standartinio nuokrypio vertė yra artima 0 (0,0232), tad tai reiškia, kad R_1 reikšmės yra tankiai išsidėsčiusios duomenų vidurkio atžvilgiu, t. y. PST tekstų kalbinė raiška šio indikatoriaus kontekste yra panaši, jie nelabai skiriasi vienas nuo kito. Variacijos koeficiento reikšmė (0,0253; žr. 25 lentelę) taip pat rodo šią tendenciją. Kvartilinio pločio vertė atskleidžia, kad R_1 reikšmių sklaida PST yra mažesnė nei AST ir MST, t. y. pagal (retesnių) žodžių ir žodžių formų įvairovės parametą PST tekstai tarpusavyje skiriasi mažiau nei AST ir MST.

Asimetrijos koeficiento vertė R_1 indikatoriumi yra neigiama (-1,630), o tai rodo didelę indikatoriaus verčių išsidėstymo asimetriją (Bulmer, 1979, 63). Toks verčių išsidėstymas reiškia, kad PST dažnesnė mažesnė (retesnių) žodžių ir jų formų įvairovė. Be to, eksceso koeficiento vertė (13,6) rodo, kad tarp PST tekstų esama nemažai tokių, kuriems būdinga netipiška – labai didelė arba labai maža – (retesnių) žodžių ir jų formų įvairovė. Į šį polinkį atkreipiamas dėmesys renkantis metodus kitam funkcinių stilių kalbinės raiškos analizės etapui.

4.3.1.4 PST tekstų RR_{mc} indikatoriaus charakteristikos

Indikatorius RR_{mc} (Kubát, Matlach ir Čech, 2014, 16) šios disertacijos tyrimo kontekste žymi tekste vartojamų žodžių ir žodžių formų koncentruotumą. Tekstui, kurio žodžių ir žodžių formų koncentruotumas yra didelis, būdinga

tai, kad jį sudaro vos keletas itin dažnai pasikartojančių žodžių ir žodžių formų, o didžioji dauguma žodžių ir žodžių formų pasikartoja žymiai rečiau. RR_{mc} vertinamo tekste vartojamų žodžių koncentruotumo galimos reikšmės pasiskirsto nuo 0 iki 1. Norint nustatyti RR_{mc} pagrindines statistines savybes PST tekstų atveju, buvo atlikta aprašomoji duomenų (indikatoriaus verčių visiems PST tekstams) analizė (Čekanavičius ir Murauskas, 2001, 25). RR_{mc} reikšmės PST pasiskirstė nuo 0,8321 (mažiausia reikšmė, žyminti mažiausių žodžių ir jų formų koncentraciją PST; ją įgijo tekstas, nepriskirtas jokiai skilčiai/teminei kategorijai (*Nenurodyta kategorija*)) iki 0,9932 (didžiausia reikšmė, žyminti didžiausių žodžių ir jų formų koncentraciją PST; ją įgijo *Sporto* skilčiai priskirtas tekstas) (plg.: AST – 0,7398 iki 0,9858 (plačiau žr. 4.1.1.4 sk.), o MST – nuo 0,8884 iki 0,9779 (plačiau žr. 4.2.1.4 sk.)). Indikatoriaus verčių PST tekstams vidurkis (0,9743) mažesnis nei mediana (0,9751), tad išskirčių (itin didelių arba mažų verčių) galima tikėtis mažų verčių pusėj. Kitaip tariant, PST tekstai, kuriems būdinga itin maža žodžių ir jų formų koncentracija, turėtų būti netipiškos išimty.

Kaip matyti iš 25 lentelės, indikatoriaus reikšmių standartinio nuokrypio vertė yra artima 0 (0,0077). Tai rodo, kad RR_{mc} reikšmės yra tankiai išsidėsčiusios apie vidurkį, t. y. tiriami publicistiniai tekstai tarpusavyje, šio indikatoriaus kontekste, mažai skiriasi. Variacijos koeficiento reikšmė yra 0,0079 (žr. 25 lentelę). Tai dar kartą rodo, kad RR_{mc} publicistinio stiliaus tekstyno tekstams būdinga maža skirtingų reikšmių įvairovė, t. y. šių tekstų žodžių ir jų formų koncentracija yra panaši, tad šiame kontekste tarpusavyje gana panašūs. Pagal kvartilinį plotį, PST tekstams būdingesnė mažesnė RR_{mc} verčių sklaida nei AST ir MST (žr. 13, 20 ir 25 lenteles), tad tai rodo, kad žodžių ir žodžių formų atžvilgiu PST tekstai tarpusavyje skiriasi mažiau nei AST ir MST tekstai.

Asimetrijos koeficiento vertė RR_{mc} indikatoriumi yra neigiama (-2,289). Tai rodo, kad PST dažnesni tekstai, kuriems būdingesnė mažesnė žodžių ir jų formų koncentracija. Eksceso koeficiento vertė (24) rodo, kad duomenims būdingas didesnis kiekis išskirčių nei tai tipiška esant normaliajai kreivei (Westfall, 2014). Kitaip sakant, pagal žodžių ir žodžių formų koncentruotumo parametą PST esama daug netipiškos raiškos atvejų, į kurių mastą svarbu atsižvelgti renkantis metodus kitam tyrimo etapui.

4.3.1.5 MATTR indikatoriaus charakteristikos PST

MATTR šioje disertacijoje laikomas informacijos srauto įverčiu (Popescu, 2009, 233-234) (plačiau žr. 1.4.2.4 sk.). Pagrindinių statistinių MATTR

savybių PST tekstuose nustatymui buvo atlikta aprašomoji duomenų analizė (Čekanavičius ir Murauskas, 2001, 36), kurios rezultatų santrauka pateikta 25-oje lentelėje. Ši analizė atskleidė, kad MATTR reikšmės PST pasiskirsčiusios nuo 26,94 (mažiausia reikšmė, mažiausio informacijos srauto arba temos plėtotės rodiklis; ji priklauso tekstui, nepriskirtam jokiai skilčiai/teminei kategorijai (*Nenurodyta kategorija*)) iki 48,97 (didžiausia reikšmė, didžiausio informacijos srauto arba temos plėtotės rodiklis; ją įgijo tekstas, priklausantis *Auto* skilčiai/teminei kategorijai) (plg.: AST – 14,76 iki 48,17 (plačiau žr. 4.1.1.5 sk.), o MST – nuo 33,44 iki 47,61 (plačiau žr. 4.2.1.5 sk.)). Tai reiškia, kad, pirmu atveju, tekstui būdingas mažas skirtingų žodžių ir jų formų skaičius, o antruoju – didelis skirtingų žodžių ir jų formų skaičius. Indikatoriaus reikšmių PST tekstams vidurkis (46,04) mažesnis nei mediana (46,22). Tai rodo, kad išskirčių, t. y. itin didelių arba mažų verčių, galima tikėtis didelių verčių pusėj. Kitaip tariant, didelis kiekis skirtingų žodžių ir jų formų (didelis informacijos srautas) PST tekstuose nėra tipiškąs.

Nedidelis, lyginant su vidurkiu, indikatoriaus verčių standartinio nuokrypio rodiklis (1,2692), reiškia, kad MATTR vertės yra išsidėsčiusios arti vidurkio, t. y. verčių įvairovė maža, todėl PST tekstai pagal šį indikatorių nedaug skiriasi tarpusavyje. Variacijos koeficiento reikšmė (0,0276) patvirtina šią tendenciją. Pagal kvartilinio pločio vertę galima spręsti, kad informacijos srauto kontekste PST tekstai tarpusavyje skiriasi mažiau nei AST ir MST (žr. 13, 20 ir 25 lenteles).

Asimetrijos koeficiento vertė MATTR indikatoriui yra neigiama (-2,561). Tai reiškia, kad mažesnės nei vidurkis vertės yra dažnesnės, t. y. PST tekstuose dažnesnis mažesnis nei vidurkis informacijos srautas, kitaip – mažesnis nei vidurkis skirtingų žodžių ir žodžių formų skaičius. Eksceso koeficiento vertė (19,3) rodo, kad tarp MATTR verčių esama daugiau išskirčių nei būdinga normaliajai kreivei (Westfall, 2014). Didelis išskirčių (netipiškų verčių) mastas rodo, kad į jas reikės atsižvelgti renkantis tyrimo metodui kitam funkcinį stilių kalbinės raiškos analizės etapui.

4.3.1.6 Teminės koncentracijos (TK) indikatoriaus charakteristikos PST

Kaip jau buvo minėta, TK indikatorius vertina, kiek teksto raiška sukoncentruota į gvildenamą temą (Kubát, Matlach ir Čech, 2014, 48; Čech, 2016; Kubát ir Čech, 2016b) (plačiau žr. 1.4.3.51 sk.). Disertacijoje tiriamame PST TK reikšmės pasiskirsčiusios intervale nuo 0 (mažiausia reikšmė – žymi teminį neutralumą (Čech, Garabík ir Altmann, 2015); šią vertę įgijo 1657 tekstai, dauguma jų priklauso *Gyvenimo*, *Piliečio* ir *Veidų* skiltims; teminis

neutralumas reiškia, kad šiuose tekstuose virš h-taško, t. y. tarp dažniausių žodžių formų, turinio/teminių žodžių formų neaptikta) iki 2,4 (didžiausia reikšmė – žymi didžiausią teminę koncentraciją; ją įgijo *Sporto* skilčiai priklausantis tekstas) (plg.: AST – 0,0505 iki 1,8833 (plačiau žr. 4.1.1.6 sk.), o MST – nuo 0 iki 0,92857 (plačiau žr. 4.2.1.6 sk.)). Siekiant ištirti bazines TK statistines savybes PST tekstams, atlikta aprašomoji duomenų analizė. Jos santrauka pateikta 25-oje lentelėje. Indikatoriaus verčių PST tekstams vidurkis (0,2586) didesnis nei mediana (0,1930), tad išskirčių, t. y. itin didelių arba mažų verčių, galima tikėtis didelių verčių pusėj. Kitaip tariant, itin temiškai koncentruoti tekstai PST turėtų būti išimtis, o ne vyraujanti tendencija.

Indikatoriaus reikšmių standartinio nuokrypio vertė, lyginant su vidurkiu, yra itin didelė (0,2470), o tai reiškia, kad TK reikšmės yra išsisklaidžiusios (Field et al., 2012, 40), t. y. pagal TK raiškos mastą, publicistinio stiliaus tekstai labai skiriasi tarpusavyje. Variacijos koeficiento reikšmė (0,9550) rodo, kad standartinis nuokrypis sudaro net 95,5 proc. aritmetinio vidurkio reikšmės. Tad galima sakyti, kad TK indikatoriaus požiūriu PST tekstams būdinga didelė skirtingų reikšmių įvairovė arba, kitaip tariant, teminės koncentracijos laipsnis labai įvairuoja, tad tekstai šiuo požiūriu labai skiriasi vienas nuo kito. Kvartilinis plotis rodo, kad PST būdinga didesnė TK verčių sklaida nei MST, bet mažesnė nei AST. Kitaip tariant, pagal teminės koncentracijos raišką PST tekstai tarpusavyje skiriasi labiau nei MST, bet mažiau nei AST (žr. 13, 20 ir 25 lenteles).

Asimetrijos koeficiento vertė TK indikatoriui yra teigiama (1,504). Tai rodo, kad PST tekstuose teminis koncentruotumas dažniau yra didesnis negu mažesnis. Eksceso koeficiento rodiklis (6,3) reiškia, kad duomenims būdingas didesnis kiekis išskirčių nei tai tipiška esant normaliajai kreivei (Westfall, 2014). Į išskirčių mastą atsižvelgiama pasirenkant metodus kitam tyrimo etapui.

4.1.2.7 Indikatorių A ir D charakteristikos PST

Kaip jau buvo šiame darbe minėta, teksto aktyvumą, vertinamą veiksmazodžių ir veiksmazodžių bei būdvardžių sumos santykiu, žymi indikatorius A, o jam priešingą deskriptyvumą – indikatorius D (Kubát, Matlach, ir Čech, 2014, 43; Kubát, 2016, 77–78) (plačiau žr. 3.1.2.5 sk.). Tyrimas atskleidė, kad A reikšmės PST tekstuose pasiskirstė nuo 0,2727 (mažiausia reikšmė, žyminti mažiausią aktyvumo raišką) iki 0,9619 (didžiausia reikšmė, žyminti didžiausią aktyvumo raišką) (plg.: AST – 0 iki

0,9658 (plačiau žr. 4.1.1.7 sk.), o MST – nuo 0,2785 iki 0,8376 (plačiau žr. 4.2.1.7 sk.)). Tuo tarpu indikatoriaus D vertės PST yra nuo 0,0381 (žymi mažiausią deskriptyvumo / aprašomumo raišką) iki 0,7273 (žymi didžiausią deskriptyvumo / aprašomumo raišką) (plg. AST – 0 iki 0,9658 (plačiau žr. 4.1.1.7 sk.), o MST – nuo 0,1624 iki 0,7215 (plačiau žr. 4.2.1.7 sk.)). Bazinių statistinių A ir D savybių ištyrimui PST kontekste taikyta aprašomoji duomenų analizė, o jos santrauka pateikta 25-oje lentelėje. Indikatoriaus A verčių PST tekstams vidurkis yra vos mažesnis nei mediana, o indikatoriaus D – vos didesnis nei mediana. Tad keletą išskirčių A atveju galima tikėtis mažų verčių pusėje, o D atveju – didelių verčių pusėj. Kitaip tariant, itin mažai dinamiškai bei itin didelės deskriptyvumo raiškos tekstai PST neturėtų vyrauti, o būtų tik išimtys.

A ir D verčių standartinio nuokrypio rodiklis (0,0768), lyginant su vidurkiu, nėra mažas, tad tai rodo, jog vertės yra išsisklaidžiusio vidurkio atžvilgiu, t. y. vertėms būdinga įvairovė, dėl ko tekstai pagal šį indikatorių gana skiriasi tarpusavyje. Variacijos koeficiento reikšmės (žr. 25 lentelę) taip pat nurodo šią tendenciją, nors pagal deskriptyvumo raiškos įvairovę PST tekstai tarpusavyje skiriasi labiau nei pagal aktyvumo / dinamiškumo. Remiantis kvartiliniu pločiu, PST būdingesnė mažesnė A ir D verčių sklaida nei AST ir MST. Kitais žodžiais tariant, pagal aktyvumo ir deskriptyvumo raišką PST tekstai tarpusavyje skirsi mažiau nei AST ir MST.

Kaip rodo 25 lentelė, asimetrijos koeficiento vertė A indikatoriui yra vidutiniškai neigiama (-0,539), o D – vidutiniškai teigiama (0,539). Tai rodo, kad PST dažnesni tekstai, kurių aktyvumo raiška mažesnė, o deskriptyvumo – didesnė. Eksceso koeficiento vertė A ir D yra 3,7 rodo, kad duomenyse išskirčių turėtų pasitaikyti daugiau nei būdinga normaliajai kreivei (Westfall, 2014). Į išskirtis (netipiškai didelės arba mažas vertės) atsižvelgiama, nes jų mastas turi įtakos renkantis metodus tolesnei analizei.

4.3.1.8 Tarpveiksmažodinių atstumų (VA) indikatoriaus charakteristikos PST tekstuose

Indikatorius VA apibūdina teksto sintaksinės struktūros kompleksumą (Kubát, 2016, 70–71), nors, žinoma toks vertinimas yra labai supaprastintas (plačiau žr. 1.4.3.3 sk.). VA reikšmės PST pasiskirsčiusios nuo 2,652 (mažiausia reikšmė, žymi sintaksinės struktūros požiūriu paprasčiausią tekstą, kuriame vidutinis atstumas tarp 2 iš eilės einančių veiksmažodžių yra 2,652 žodžiai ir žodžių formos; ji priklauso *Piliečio* skilties tekstui) iki 47,333 (didžiausia reikšmė, žymi sintaksinės struktūros požiūriu sudėtingiausią

tekstą, kuriame vidutinis atstumas tarp 2 iš eilės einančių veiksmažodžių yra 47,333 žodžiai ir žodžių formos; ji priklauso tekstui, nepriskirtam jokiai skilčiai/teminei kategorijai (*Nenurodyta kategorija*)) (plg.: AST – nuo 3,098 iki 622,2 (plačiau žr. 4.1.1.8 sk.), o MST – nuo 3,233 iki 22,500 (plačiau žr. 4.2.1.8 sk.)). Siekiant ištirti VA bazines statistines savybes PST kontekste, atlikta aprašomoji duomenų analizė, kurios santrauka pateikta 25-oje lentelėje. Indikatoriaus verčių tiriamiems publicistiniams tekstams vidurkis (4,661) yra didesnis nei mediana (4,489), todėl išskirčių galima tikėtis didelių verčių pusėj. Kitaip tariant, labai sudėtingos sintaksinės struktūros tekstai (t. y. tokie, kuriose vidutinis atstumas tarp dviejų iš eilės einančių veiksmažodžių bus labai didelis) publicistinio stiliaus tekstų atžvilgiu, tikėtina, priklausys išimtims.

Indikatoriaus verčių standartinio nuokrypio vertė (1,1784), lyginant su vidurkiu yra didelė, tad VA vertės yra išsisklaidžiusios to paties vidurkio atžvilgiu (Field et al., 2012, 40). Kitaip tariant, PST tekstams būdinga sintaksės kompleksiško įvairovė, todėl šio tekstyno tekstai gana skiriasi tarpusavyje. Variacijos koeficiento reikšmė (žr. 25 lentelę) taip pat rodo šią tendenciją. Pagal kvartilinio pločio vertę galima sakyti, kad PST tekstai teksto sintaksinės struktūros kompleksiško kontekste skiriasi vos mažiau nei MST ir gerokai mažiau nei AST ir tekstai (žr. 13, 20 ir 25 lenteles).

Asimetrijos koeficiento vertė VA indikatoriui yra teigiama (9,433) Kitaip sakant, duomenyse žemesnės vertės dažnesnės nei didesnės arba PST tekstuose atstumai tarp dviejų iš eilės einančių veiksmažodžių dažniau yra mažesni negu didesni, t. y. dažnesni yra sintaksės požiūriu paprastesni tekstai. Eksceso koeficiento vertė (252,3) rodo, kad duomenyse tikėtina rasti daug išskirčių, t. y. netipiškai didelių ar mažų verčių (Westfall, 2014). Į tai atsižvelgiama renkantis tolesnius tyrimo metodus.

4.3.2 Indikatorių koreliacija PST tekstams

Šiame poskyryje pristatoma tiriamų indikatorių koreliacinė analizė, skirta statistinio ryšio tarp indikatorių poros nustatymui ir įvertinimui. Tam pasirinktas Spirmeno (Spearman) koreliacijos koeficientas (Field, Miles ir Field, 2012, 223), nes PST kalbinę raišką žyminčių indikatorių vertėms, lygiai kaip ir AST bei MST atvejais, buvo būdingos išskirtys (Mukaka, 2012).

4.3.2.1 Išskirtys PST duomenyse

Kadangi išskirtys (netipiškos duomenų vertės – itin didelės arba itin mažos) yra svarbios tolesnių FS kalbinės raiškos požymių analizės metodų pasirinkimui, šiame skyriuje pristatoma išskirčių PST analizė. 26-oje lentelėje galima matyti, visų tirtų indikatorių vertėms būdingos išskirtys, kitaip – itin didelės arba itin didelės vertės, išsiskiriančios bendrame duomenų kontekste. Šių netipiškų verčių PST aptikta panašiai kaip ir AST (žr. 16 lentelę), bet gerokai daugiau nei MST (žr. 21 lentelę). Išskirčių intervalai indikatoriams buvo apskaičiuoti remiantis kvartilinio pločio (KP) statistika (plačiau žr. 3.1 sk.).

26 lentelė. PST: indikatorių verčių išskirtys

Indikatorius	Išskirtys
VŽFI	Vertės, mažesnės nei 5,3 (rasta 47) ir didesnė nei 7,4 (rasta 43)
<i>a</i>	Vertės, mažesnės nei 5,4 (rasta 12) ir didesnės nei 23 (rasta 332)
R ₁	Vertės, mažesnės nei 0,86 (rasta 260) ir didesnės nei 0,97 (rasta 3)
RR _{mc}	Vertės, mažesnės nei c (rasta 244) ir didesnės nei 0,99 (rasta 100)
MATTR	Vertės, mažesnės nei 44 (rasta 447) ir didesnės nei 49 (rasta 5)
TK	Vertės, didesnės nei 0,85 (rasta 405)
A	Vertės, mažesnės nei 0,53 (rasta 175) ir didesnės nei 0,94 (rasta 5)
D	Vertės, mažesnės nei 0,062 (rasta 5) ir didesnės nei 0,47 (rasta 175)
VA	Vertės, didesnės nei 6,6 (rasta 486)

Tad, kaip rodoma 26-oje lentelėje, indikatoriaus VŽFI vertės, mažesnės nei 5,3 (priklauso 47 PST tekstams), bet didesnės nei 7,4 (priklauso 43 PST tekstams), laikomos išskirtimis. Kitaip tariant, vidutiniškai trumpesni nei 5,3 simboliai ir vidutiniškai ilgesni nei 7,4 simbolių žodžiai ir žodžių formos yra netipiški PST tekstams. Taip pat *a* vertės, mažesnės nei 5,4 (tokių PST rasta 12, t. y. dvylikos tekstų *a* vertės šiame tekстыne yra mažesnės nei 5,4), bet didesnės nei 23 (rasti 332 PST tekstai, turintys tokias vertes) priklauso išskirtims. Tai reiškia, kad labai dažnų žodžių ir žodžių formų sritys, kurių didumo įvertis yra mažesnis nei 5,4, bet didesni nei 23 yra netipiški atvejai PST, galintys iškreipti tyrimo rezultatus, jei naudojami standartiniai statistikos metodai.

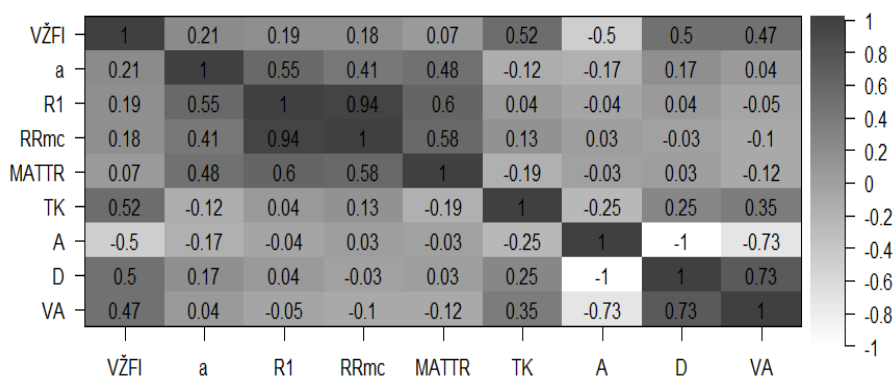
Indikatoriaus R_1 , žyminčio (retesnių) žodžių ir žodžių formų įvairovę, kontekste išskirtys yra vertės, mažesnės nei 0,86 ir didesnės nei 0,97. Kitaip sakant, PST tekstuose žyminčio (retesnių) žodžių ir žodžių formų įvairovė, kurios įvertis mažesnis nei 0,86 ir didesnis nei 0,97 laikoma netipiška. O netipišką žodžių ir žodžių formų koncentruotumą PST rodo RR_{mc} vertė, kurios yra mažesnės nei 0,97 ir didesnės nei 0,99. PST nebūdingas informacijos srautas yra toks, kurio įvertis yra mažesnis nei 44 ir didesnis nei 49. Be to, publicistiniams tekstams netipiška yra itin didelė teminė koncentracija, t. y. tokia, kurios įvertis didesnis nei 0,85. Dar publicistiniams tekstams nebūdingas, t. y. sudaro išimtis, itin didelė arba maža aktyvumo / dinamiškumo raiška (indikatoriaus A vertės mažesnės nei 0,53 ir didesnės nei 0,99) bei itin maža bei didelė deskriptyvumo / aprašomumo raiška (indikatoriaus D vertės, mažesnės nei 0,062 ir didesnės nei 0,47. Galiausiai PST netipiški yra tekstai, kurių sintaksinė struktūra yra itin kompleksiška (indikatorius VA), t. y. tokie tekstai, kuriuose vidutinis atstumas tarp 2 iš eilės einančių yra didesnis nei 6,6 žodžiai ir žodžių formos.

Kadangi PST kontekste indikatorių vertėms būdingos išskirtys (netipiškai didelės arba mažos vertės), ryšiams tarp indikatorių nustatyti ir įvertinti taikytas ne įprastas Pirsono, bet ranginis Spirmeno koreliacijos koeficientas (Mukaka, 2012; Lapinskas, 2003, 159).

4.3.2.2 Koreliacinės analizės rezultatai PST

PST indikatorių koreliacinės analizės rezultatai pateikti 11-ame paveiksle. Kaip jau buvo šiame darbe minėta, kuo didesnė koreliacijos koeficiento reikšmė (t. y. arčiau 1 arba -1) – tuo ryšys stipresnis. Kuo šviesesne spalva 11-ame paveiksle nuspalvintas kvadratis, tuo stipresnis neigiamas ryšys sieja 2 indikatorius (vieno jų reikšmėms didėjant, kito – mažėja ir atvirkščiai). Lygiai taip pat, kuo tamsesne spalva nuspalvintas kvadratis, tuo stipresnis teigiamas ryšys sieja 2 indikatorius (vieno jų reikšmėms didėjant/mažėjant, kito taip pat didėja/mažėja).

Indikatorių koreliacijos diagrama



11 pav. Koreliacinės analizės diagrama PST

11-ame paveiksle pateiktoje diagramoje galima matyti, kad labai stipri koreliacija nustatyta 2 indikatorių poroms – RR_{mc} ir R_1 bei A ir D ¹¹⁰. Vidutiniškai stiprus ryšys sieja 11 indikatorių porų: TK ir $VŽFI$; A ir $VŽFI$; D ir $VŽFI$; VA ir $VŽFI$; R_1 ir a ; RR_{mc} ir a ; $MATTR$ ir a ; $MATTR$ ir R_1 ; $MATTR$ ir RR_{mc} ; VA ir A ; VA ir D . Silpna koreliacija rasta tarp 4 indikatorių poroms: $VŽFI$ ir a ; A ir TK ; D ir TK ; VA ir TK . O labai silpna – 19 indikatorių porų: R_1 ir $VŽFI$; RR_{mc} ir $VŽFI$; $MATTR$ ir $VŽFI$; D ir a ; VA ir a ; TK ir R_1 ; D ir R_1 ; TK ir RR_{mc} ; A ir RR_{mc} ; D ir $MATTR$; TK ir a ; A ir a ; A ir R_1 ; VA ir R_1 ; D ir RR_{mc} ; VA ir RR_{mc} ; TK ir $MATTR$; A ir $MATTR$; VA ir $MATTR$. 27-oje lentelėje pristatomi ryšių tarp indikatorių porų statistinio reikšmingumo įverčiai.

¹¹⁰ Kaip galima matyti, A ir D yra aiškiai priklausomi kintamieji, tad vieno jų būtų galima atsisakyti, tačiau, siekiant susidaryti bendrą vaizdą ir sekant čekų mokslininkų darbais (pavyzdžiui, Kubát (2015)), šiame tyrimo etape pasiliekami abu indikatoriai. FS tarpusavio palyginimui taikant neparametrinius išvadų statistikos testus (žr. 5 sk.), iš šių dviejų indikatorių pasilieka tik indikatorius A .

27 lentelė. PST: koreliacijos statistinis reikšmingumas (p -vertės)

	VŽFI	a	R_1	RR_{mc}	MATTR	TK	A	D	VA
VŽFI		0,00 0	0,00 0	0,000	0,000	0,00 0	0,000	0,000	0,00 0
a	0,000		0,00 0	0,000	0,000	0,00 0	0,000	0,000	0,00 0
R_1	0,000	0,00 0		0,000	0,000	0,00 0	0,000	0,000	0,00 0
RR_{mc}	0,000	0,00 0	0,00 0		0,000	0,00 0	0,003 2	0,003 2	0,00 0
MAT TR	0,000	0,00 0	0,00 0	0,000		0,00 0	0,000 4	0,000 4	0,00 0
TK	0,000	0,00 0	0,00 0	0,000	0,000		0,000	0,000	0,00 0
A	0,000	0,00 0	0,00 0	0,003 2	0,0004	0,00 0		0,000	0,00 0
D	0,000	0,00 0	0,00 0	0,003 2	0,0004	0,00 0	0,000		0,00 0
VA	0,000	0,00 0	0,00 0	0,000	0,000	0,00 0	0,000	0,000	

Kaip galima matyti 27 lentelėje, koreliacija buvo statistiškai reikšminga visoms indikatorių poroms, t. y. visoms indikatorių poroms $p \leq 0,05$. Poskyriuose 4.3.2.2.1-4.3.2.2.3 pateikiami detalūs koreliacinės analizės PST rezultatai.

4.3.2.2.1 Labai stipri koreliacija tarp indikatorių

Duomenų analizė parodė, kad labai stipri statistiškai reikšminga koreliacija (Čekanavičius ir Murauskas, 2001, 126) randama tarp indikatorių RR_{mc} ir R_1 (0,94; žymi labai stiprią teigiamą koreliaciją) bei A ir D (-1; žymi labai stiprią neigiamą koreliaciją). Lyginant su kitų tiriamųjų tekstynų rezultatais, AST atveju ryšio stiprumas tarp RR_{mc} ir R_1 yra toks pats, kaip ir PST (žr. 5 paveikslą), o MST atveju – mažesnis nei PST (žr. 8 paveikslą). Indikatorius RR_{mc} leidžia įvertinti PST tekstų žodžių ir žodžių formų koncentruotumą, o R_1 – (retesnių) žodžių ir žodžių formų įvairovę. Tarp RR_{mc} ir R_1 matome labai stiprų teigiamą ryšį, rodantį, kad didėjant/mažėjant žodžių ir žodžių formų koncentruotumui, kartu didėja/mažėja ir (retesnių) žodžių ir žodžių formų. RR_{mc} apima visus tekstą sudarančius žodžius, o R_1 – tik dalį, t. y. retesnius, dažninėje lentelėje esančius žemiau h-taško (plačiau žr. 1.4.2.1 sk.). Tad gali nutikti, kad turėsime tekstą, kuriame keletas žodžių pasikartoja itin daug kartų, o likusi dauguma – vos po porą ar tik vieną. Tokiu būdu ir gauname tiek didelį / mažą žodžių ir žodžių formų koncentruotumą, tiek didelę / mažą (retesnių)

žodžių ir žodžių formų įvairovę. Indikatoriai A ir D žymi teksto „toną“ per aktyvumo (A) ir deskriptyvumo (D) dimensijas. Šie indikatoriai yra vienas kitam priešingi, tad juos sieja itin stiprus neigiamas statistiškai reikšmingas ryšys, rodantis, kad tekste stiprėjant aktyvumo raiškai, deskriptyvumo raiška mažėja ir atvirkščiai.

4.3.2.2.2 Vidutiniškai stipri koreliacija tarp indikatorių

Vidutinio stiprumo statistiškai reikšminga koreliacija (Čekanavičius ir Murauskas, 2001, 126) nustatyta 11 indikatorių porų: TK ir VŽFI; A ir VŽFI; D ir VŽFI; VA ir VŽFI; R_1 ir a ; RR_{mc} ir a ; MATTR ir a ; MATTR ir R_1 ; MATTR ir RR_{mc} ; VA ir A; VA ir D (žr. 11 pav.). 9 indikatorių poras sieja teigiamas ryšys (TK ir VŽFI; D ir VŽFI; VA ir VŽFI; R_1 ir a ; RR_{mc} ir a ; MATTR ir a ; MATTR ir R_1 ; MATTR ir RR_{mc} ; VA ir D), t. y. vieno indikatoriaus vertėms didėjant/mažėjant, kito taip pat didėja / mažėja, o 2 – neigiamas (A ir VŽFI; VA ir A), t. y. vieno indikatoriaus vertėms didėjant, kito – mažėja ir atvirkščiai. Toliau plačiau apžvelgiami vidutinio stiprumo ryšiai, nustatyti tarp šių indikatorių porų.

Vidutinio stiprumo teigiamas ryšys tarp TK ir VŽFI rodo, kad didėjant arba mažėjant teminei koncentracijai (TK), teksto sudėtingumas/suprantamumas (VŽFI) taip pat didėja arba mažėja, t. y. ilgėja/trumpėja vidutinis žodžių ir jų formų ilgis. A ir VŽFI siejantis neigiamas ryšys reiškia, kad didėjant teksto aktyvumo (A) raiškai (kitaip – daugėjant tekste veiksmažodžių, lyginant su būdvardžiais), jo sudėtingumas (VŽFI) mažėja (kitaip – vidutinis žodžių ir jų formų ilgis trumpėja) ir atvirkščiai. D ir VŽFI sieja teigiamas ryšys, kuris parodo, jog didėjant/mažėjant teksto deskriptyvumo (D) raiškai (t. y. daugėjant/mažėjant tekste būdvardžių, lyginant su veiksmažodžiais), jo sudėtingumas/suprantamumas (VŽFI) taip pat didėja/mažėja (t. y. ilgėja/trumpėja vidutinis žodžių ir žodžių formų ilgis). Be to, teigiamas ryšys tarp VA ir VŽFI nusako, kad kuo sudėtingesnė/paprastesnė teksto sintaksinė struktūra (t. y. didesnis/mažesnis atstumas tarp 2 iš eilės einančių veiksmažodžių), tuo šis tekstas yra sudėtingesnis (sunkiau suprantamas/paprastesnis (lengviau suprantamas), t. y. tuo jam būdingesni vidutiniškai ilgesni/trumpesni žodžiai ir žodžių formos.

Nustatytas teigiamas ryšys tarp indikatorių R_1 ir a rodo, kad kuo didesnė/mažesnė tiriamų publicistinių tekstų (retesnių) žodžių ir jų formų įvairovė, tuo mažesnė / didesnė jų labai dažnų žodžių sritis dažninėje žodžių ir žodžių formų lentelėje. Lygiai taip pat, ryšys (teigiamas) tarp RR_{mc} ir a nusako, kad kuo didesnis/mažesnis žodžių ir jų formų koncentruotumas tekste,

tuo mažesnė / didesnė jame labai dažnų žodžių sritis dažninėje žodžių ir žodžių formų lentelėje. Panašiai teigiamas ryšys tarp MATTR ir a nurodo, kad kuo didesnis/mažesnis informacijos srautas tekste (kitaip – kuo daugiau / mažiau tekste skirtingų žodžių ir žodžių formų), tuo tekstui būdingesnė mažesnė / didesnė gausa labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų.

Teigiamas ryšys tarp MATTR ir R_1 rodo, kad kuo didesnis/mažesnis informacijos srautas tekste (kitaip sakant, kuo daugiau tekste skirtingų žodžių ir žodžių formų), tuo jo (retesnių) žodžių ir žodžių formų įvairovė yra didesnė/mažesnė. Be to, ryšys (teigiamas) tarp MATTR ir RR_{mc} žymi, kad kuo didesnis/mažesnis informacijos srautas būdingas tekstui (t. y. kuo daugiau tekste skirtingų žodžių ir žodžių formų), tuo tipiškėsnis jam didesnis/mažesnis žodžių ir žodžių formų koncentruotumas. Neigiamas ryšys tarp VA ir A nurodo, kad kuo PST tekstų sintaksinė struktūra sudėtingesnė (kitaip tariant, kuo šiuose tekstuose atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis), tuo mažesnė juose aktyvumo raiška, t. y. tuo mažiau juose, lyginant su būdvardžiais, yra veiksmažodžių bei atvirkščiai. Galiausiai vidutinio stiprumo teigiamas ryšys tarp VA ir D pažymi, kad kuo teksto sintaksinė struktūra sudėtingesnė/paprastesnė (kitaip tariant, kuo tekste atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis/mažesnis), tuo jame deskriptyvumo raiška didesnė/mažesnė (t. y. tuo daugiau/mažiau tekste būdvardžių, lyginant su veiksmažodžiais).

4.3.2.2.3 Silpna ir labai silpna koreliacija tarp indikatorių

Silpną koreliaciją žymi Spirmeno koreliacijos koeficiento reikšmės $[-0,4; -0,2]$ arba $[0,2; 0,4]$ (Čekanavičius ir Murauskas, 2001, 126). Ji nustatyta 4 indikatorių poroms: VŽFI ir a ; A ir TK; D ir TK; VA ir TK. Iš jų teigiamas ryšys (vienos indikatorius vertėms didėjant/mažėjant, kito irgi didėja/mažėja) rastas 3 indikatorių poroms (VŽFI ir a ; D ir TK; VA ir TK), neigiamas – 1 porai (A ir TK) Labai silpną koreliaciją žymi Spirmeno koreliacijos koeficiento reikšmės $[-0,2; 0]$ arba $[0; 0,2]$ (Čekanavičius ir Murauskas, 2001, 126). Ji nustatyta 19 indikatorių porų, iš jų 10 porų sieja teigiamas ryšys (R_1 ir VŽFI; RR_{mc} ir VŽFI; MATTR ir VŽFI; D ir a ; VA ir a ; TK ir R_1 ; D ir R_1 ; TK ir RR_{mc} ; A ir RR_{mc} ; D ir MATTR), o 9 – neigiamas (TK ir a ; A ir a ; A ir R_1 ; VA ir R_1 ; D ir RR_{mc} ; VA ir RR_{mc} ; TK ir MATTR; A ir MATTR; VA ir MATTR).

Koreliacija tarp indikatorių PST kontekste buvo statistiškai reikšmingai visais atvejais, nesvarbu, kokio stiprumo ryšys buvo nustatytas. Kad visi ryšiai

buvo statistiškai reikšmingi, galima matyti 27-oje lentelėje, kur visos p vertės yra mažesnės nei 0,05 (ryšys laikomas statistiškai reikšmingu, jei $p \leq 0,05$ (Field, Miles ir Field, 2012, 51—52)). Tad galima sakyti, kad PST tekstų atžvilgiu visi disertacijoje naudojami kalbinę raišką žymintys indikatoriai yra stipriau ar silpniau susiję, t. y. turi įtakos vienas kito verčių kitimui.

4.3.3 PST pagrindinių komponentių analizės (PKA) rezultatai

4.3.3.1 Indikatorių svarba reprezentuojant publicistinio stiliaus kalbinę raišką

Pagrindinių komponentių analizė (PKA) leido įvertinti kiekvieno iš 9 indikatorių, naudotų tyrime (a , VŽFI, R_1 , RR_{mc} , VA, TK, MATTR, A ir D), indėlį ir svarbą apibūdinant publicistinį stilių bei išskirti geriausiai šį stilių apibūdinančių indikatorių kombinaciją. 28-oje lentelėje pateikiama PST kalbinę raišką reprezentuojančių indikatorių PKA rezultatų santrauka (lentelėje „PK1“, ..., „PK9“ žymimos pagrindinės komponentės). Kaip jau minėta, atliekant analizę, vertės pasukamos, ieškant krypties, kurioje duomenų sklaida yra didžiausia. Toji kryptis vadinama pirmąja pagrindine komponente. Pagrindinės komponentės yra linijinė kintamųjų (kalbinę raišką žyminčių indikatorių) iš pirminio, „nepasuktų“ duomenų kombinacija. Pirmoji pagrindinė komponentė apibūdina didžiausią duomenų verčių variaciją (žr. 28-ą lentelę, eilutę „Variacijos proporcija, %“). Kuo didesnę duomenų variantiškumo dalį ši komponentė apibūdina, tuo daugiau informacijos apie tiriamą objektą (PST kalbinės raiškos požymius) ji išreiškia. Antroji pagrindinė komponentė randama tokiu pat būdu kaip ir pirmoji, tik dar pridėdant sąlygą, kad komponentės viena su kita nekoreliuotų, t. y. būtų viena kitai statmenos (Everitt ir Hothorn, 2011, 63—65). Antroji komponentė paaiškina didžiausią dalį likusios dispersijos, t. y. iš to, kas liko nuo pirmosios pagrindinės komponentės. Pagrindinių komponentių yra tiek, kiek kintamųjų (9, pagal tiriamų indikatorių skaičių; žr. 28-ą lentelę).

28-oje lentelėje galima matyti, kad pirmos trys pagrindinės komponentės paaiškina didžiąją dalį visos duomenų dispersijos. Pirmoji pagrindinė komponentė PK1 paaiškina 35,503 proc. duomenų variacijos, antroji pagrindinė komponentė PK2 – 31,609 proc. duomenų variacijos, o trečioji – 12,419 proc. duomenų variacijos (žr. 28-os lentelės eilutę „Variacijos proporcija, %“). Kartu jos apibūdina 79,530 proc. visos duomenų variacijos ($PK1 + PK2 + PK3 = 35,503 + 31,609 + 12,419 = 79,530$; žr. 28-os lentelės eilutę „Sukauptinė variacijos proporcija, %“). Likusios 6 komponentės

kiekviena atskirai paaiškina vis mažesnę dalį duomenų variacijos (žr. 28-os lentelės eilutę „Variacijos proporcija, %“). Kartu jos apibūdina tik 20,47 proc. indikatorių reikšmių ($PK4 + PK5 + PK6 + PK7 + PK8 + PK9 = 6,765 + 5,520 + 4,303 + 3,316 + 0,67 + 0,00 = 20,47$; žr. 28-os lentelės eilutę „Variacijos proporcija, %“).

Visų kalbinę raišką žyminčių indikatorių reikšmių atvaizdavimui pakanka 7 komponentių (sudėjus pirmas 7 komponentes, jos visos kartu paaiškina 99,433 proc. visos reikšmių variacijos). Tad, kaip įprasta atliekant PKA, tolesnei analizei pasiliekomos tik didžiausią indikatorių reikšmių variacijos dalį paaiškinančios komponentės PK1, PK2 ir PK3. Jų tikrinių vektorių reikšmės (rodo, kokią dalį pradinių duomenų variacijos paaiškina komponentės (Janilionis, Morkevičius ir Rauleckas 2008, 171); žr. 28 lentelę) yra didesnė nei 1, tad tai reiškia, jog komponentė, turinti šią reikšmę, reprezentuoja didesnę dalį indikatorių reikšmių variacijos nei vienas kuris indikatorius iš pirminio „nepasukto“ indikatorių verčių rinkinio (Kassambara, 2017, 18).

28 lentelė. PKA rezultatų santrauka PST: duomenų variacijos proporcijos, kurias paaiškina komponentės

	PK1	PK2	PK3	PK4	PK5	PK6	PK7	PK8	PK9
Tikrinių vektorių reikšmės	3,195	2,845	1,118	0,609	0,497	0,387	0,298	0,051	0,000
Variacijos proporcija (%)	35,50 3	31,60 9	12,41 9	6,765	5,520	4,303	3,316	0,67	0,000
Sukauptinė variacijos proporcija (%)	35,50 3	67,11 2	79,53 0	86,29 5	91,81 5	96,11 8	99,43 3	100,0 00	100,0 00

Remiantis PKA rezultatais (28 lentelė), toliau buvo analizuojama, kurie indikatoriai tiriant PST kalbinę raišką reprezentuoja didžiausią dalį duomenų reikšmių, t. y. yra svarbiausi arba, kitaip sakant, geriausiai apibūdina šio stiliaus raišką. Kartu buvo siekiama nustatyti kiekvieno indikatoriaus indėlių charakterizuojant šį FS. Atskirų indikatorių svarba įvertinta pagal kintamųjų dvimatėje PKA rezultatų diagramoje, kurioje kiekvienas duomenų taškas turi dvi koordinates – horizontalioje ašyje ir vertikalioje ašyje, įgyjamas kvadratinės kosinuso vertes (\cos^2) (Husson, Josse, Le ir Mazet, 2018), t. y. kampo tarp koordinačių didumą (plačiau žr. 4.1.3.1 sk.). Kuo didesnė ši vertė,

tuo indikatorius konkrečioje komponentėje yra svarbesnis, t. y. tuo jis komponentėje geriau reprezentuojamas.

Kiekvieno indikatoriaus indėlis atskirose komponentėse skaičiuojamas procentais įvertinant indikatoriaus svarbą visoje komponentėje (Kassambara, 2017, 20) (plačiau žr. 4.1.3.1 sk.). Kuo ši procentinė išraiška didesnė, tuo indikatoriaus indėlis konkrečioje komponentėje yra didesnis. Primintina, kad pagrindinės komponentės yra indikatorių verčių kombinacijos, t. y. atlikus duomenų transformaciją, kitaip – juos „pasukus“ didžiausios dispersijos kryptimi, vienoje ašyje yra nebe vieno indikatoriaus vertės, o jų mišinys skirtingomis proporcijomis. Todėl skirtingose komponentėse indikatorių svarba bei indėlis apibūdinant PST tekstus skiriasi.

29 lentelė. PKA rezultatų santrauka PST: indikatorių komponentėse PK1, PK2 ir PK3 svarba ir indėlis

	PK1		PK2		PK3	
	Indėlis (%)	Svarba (\cos^2)	Indėlis (%)	Svarba (\cos^2)	Indėlis(%)	Svarba (\cos^2)
VŽFI	8,278	0,265	10,777	0,307	8,173	0,091
<i>a</i>	1,474	0,047	15,736	0,448	8,296	0,093
R_1	7,313	0,234	22,612	0,643	2,328	0,026
RR_{mc}	8,110	0,259	18,920	0,538	7,913	0,088
MATTR	10,628	0,340	13,080	0,372	3,596	0,040
TK	8,114	0,259	0,951	0,027	51,281	0,573
A	18,574	0,593	8,386	0,239	9,206	0,103
D	18,574	0,593	8,386	0,239	9,206	0,103
VA	18,934	0,605	1,151	0,033	0,000	0,000

Kaip galima matyti 29-oje lentelėje, PK1 geriausiai reprezentuojami yra VA, A ir D indikatoriai (jų \cos^2 vertės yra didžiausios). VA žymi teksto sintaksinės struktūros kompleksiskumą, o A ir D įvertina tekstą per aktyvumo ir deskriptyvumo dimensijas, t. y. įvertina teksto „toną“. Šių indikatorių indėlis šioje komponentėje taip pat yra didžiausias (žr. 29-os lentelę). VA, A ir D kartu apibūdina 56,082 proc. PK1 duomenų variacijos ($VA(18,934) + A(18,574) + D(18,574) = 56,082$; žr. 29 lentelę). PK1 reprezentuoja 35,503 proc. visų indikatorių duomenų verčių sklaidos. Tad PK1 kombinacija VA, A ir D nusako 19,91 proc. visų indikatorių verčių variacijos. Kitų indikatorių svarba ir indėlis šioje komponentėje yra mažesni, o mažiausiai svarbus (kartu mažiausiai prisidedantis apibūdinant PST tekstus) yra indikatorius *a*, žymintis

ribą tarp labai dažnų ir retesnių žodžių ir jų formų (jo \cos^2 vertė ir procentinis indėlis, lyginant su kitais indikatoriais, yra mažiausias). Taigi VA, A ir D komponentėje PK1 yra svarbiausi, t. y. reprezentuoja daugiausiai PK1 esančių indikatorių verčių, o indikatorius a šioje komponentėje yra mažiausiai svarbu, nes nusako tik labai mažą dalį visų PK1 esančių indikatorių verčių.

Komponentė PK2 reprezentuoja 31,609 proc. visų indikatorių reikšmių (žr. 28 lentelę). Pagal svarbą ir indėlį šioje komponentėje išsiskiria indikatoriai R_1 ir RR_{mc} , turintys didžiausias \cos^2 vertes (žr. 29 lentelę). R_1 yra žodžių ir žodžių formų įvairovės rodiklis, o RR_{mc} – žodžių ir jų formų koncentracijos indikatorius. Abu kartu šie indikatoriai apibūdina 41,53 proc. PK2 indikatorių verčių variacijos ($R_1(22,612) + RR_{mc}(18,920) = 41,53$; žr. 28 lentelę). Kadangi PK2 reprezentuoja 31,609 proc. visų indikatorių reikšmių, tai R_1 ir RR_{mc} kombinacija nusako 13,13 proc. visos indikatorių reikšmių sklaidos. Kitų indikatorių svarba ir indėlis mažesni. Mažiausiai apibūdinant PST komponentėje PK2 prisideda indikatoriai VA ir TK, turintys mažiausias \cos^2 reikšmes. Tad indikatoriai R_1 (žymi (retesnių) žodžių ir žodžių formų įvairovę tekste) ir RR_{mc} (žymi žodžių ir žodžių formų koncentruotumą tekste) yra svarbiausi PK2 kontekste apibūdinant publicistinio stiliaus tekstyno tekstus, o VA (žymi teksto sintaksinės struktūros kompleksškumą) ir TK (teksto teminės koncentracijos įvertis) – mažiausiai svarbūs.

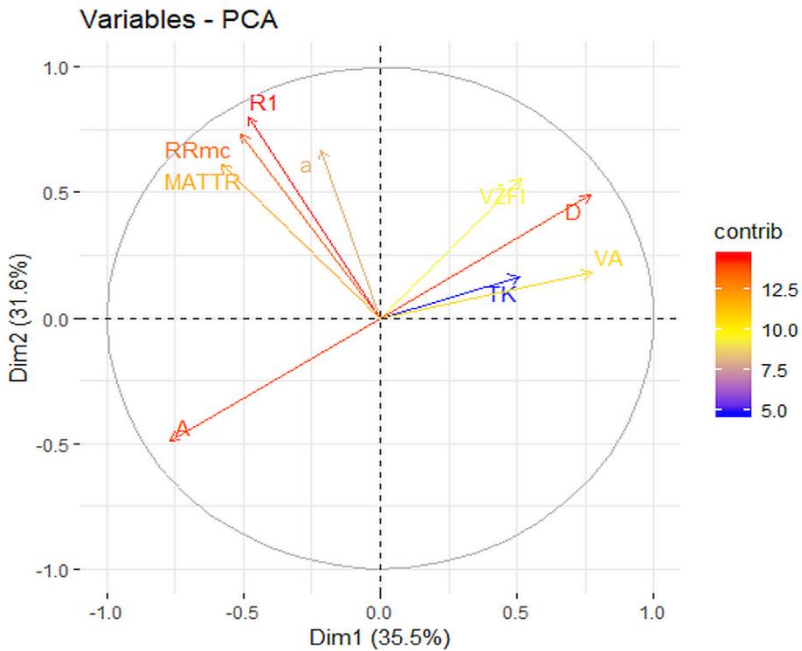
Paskutinėje analizuotoje pagrindinėje komponentėje PK3 išsiskiria vienas indikatorius – TK, apibūdinantis teksto teminę koncentraciją. Šio indikatorius svarba ir indėlis labai skiriasi nuo kitų šios komponentės indikatorių: šis indikatorius reprezentuoja net 51,281 proc. visų PK3 kintamųjų reikšmių. PK3 reprezentuoja 12,419 proc. visos indikatorių verčių sklaidos, tai TK nusako 1,6 proc. visos indikatorių verčių variacijos. Kitų indikatorių svarba ir indėlis mažesni. Mažiausias indėlis ir svarba apibūdinant publicistinį stilių komponentėje PK3 priklauso indikatoriumi VA, kurio \cos^2 reikšmė lygi 0. Tad indikatorius TK yra svarbiausias PK3 kontekste apibūdinant PST tekstus, o VA – mažiausiai svarbus.

4.3.3.2 Indikatorių indėlis reprezentuojant publicistinio stiliaus kalbinę raišką

Ištyrus atskirų indikatorių svarbą charakterizuojant publicistinio funkcinio stiliaus kalbinę raišką, atlikta ir indikatorių indėlio apibūdinant šią raišką analizė, kurios rezultatai pateikti 12-ame ir 13-ame paveiksluose. Kuo indikatorius reprezentuojančių rodyklių spalva raudonesnė, tuo jų indėlis apibūdinant PST kalbinę raišką yra didesnis. Ir kuo indikatorius reprezentuojančių rodyklių spalva mėlynesnė/labiau violetinė, tuo jų indėlis

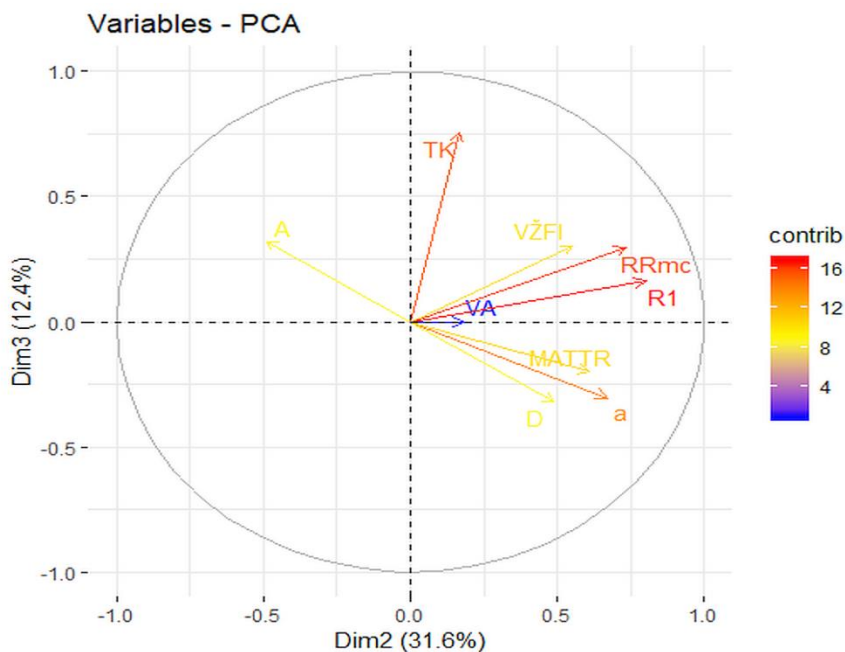
apibūdinant PST kalbinę yra mažesnis. Be to, kuo šių rodyklių smaigaliai yra arčiau paveikslė pavaizduoto apskritimo perimetro, tuo indikatorių indėlis taip pat didesnis, o kuo smaigaliai yra arčiau paveikslė pavaizduoto apskritimo centro, tuo indikatorių indėlis yra mažesnis.

12-ame paveiksle pavaizduoti atskirų indikatorių indėliai apibūdinant publicistinio funkcinio stiliaus tekstyno tekstų kalbinę raišką bendrame pirmųjų dviejų pagrindinių komponentų PK1 ir PK2 kontekste. Vėl horizontalioje ašyje atvaizduota PK1 (ši komponentė reprezentuoja didžiausią dalį indikatorių reikšmių variacijos), o vertikalioje ašyje – PK2 (ši komponentė reprezentuoja didžiausią dalį indikatorių reikšmių variacijos, likusios po PK1). Tad galima matyti, kad didžiausias indėlis apibūdinant PST tekstus PK1 ir PK2 kontekste priklauso indikatoriams R_1 , RR_{mc} , A ir D (12 pav.) – juos vaizduojančių rodyklių spalva yra raudoniausia, o tų pačių rodyklių smaigaliai baigiasi arčiausiai paveikslė pavaizduoto apskritimo perimetro. Kitaip sakant, PK1 ir PK2, reprezentuojančiose daugiau nei pusę PST kalbinę raišką apibūdinančių indikatorių reikšmių variacijos (28 lentelė), žodžių formų įvairovė (R_1), žodžių formų koncentruotumas (RR_{mc}) bei „tonas“ (aktyvumo-deskriptyvumo dimensijos, t. y. indikatoriai A ir D) labiausiai prisideda charakterizuojant šį funkcinį stilių. MATTR ir VA indėlis yra kiek mažesnis (rodyklės, reprezentuojančios šį indikatorių, spalva yra labiau oranžinė nei raudona, o smaigalys baigiasi toliau nuo paveikslė pavaizduoto apskritimo perimetro nei R_1 , RR_{mc} , A ir D).



12 pav. Indikatorių indėlis PST: PK1 ir PK2

Indikatorių svarba pagal jų indėlį reprezentuojant publicistinę funkcinių stilių pagrindinėje komponentėje PK3 pavaizduota 13-ame paveiksle. Kaip ir indikatorių svarbos analizės atveju, taip ir tiriant indikatorių indėlį reprezentuojant PST, siekiant kaip galima tiksliau atvaizduoti kintamųjų reikšmes antrąja ašimi dvimatei diagramai pasirenkama antroji pagrindinė komponentė PK2, jau aptarta aukščiau. Tad 13-ame paveiksle X (horizontalioje) ašyje pavaizduota PK2 (ji reprezentuoja didesnę procentą visų kintamųjų reikšmių), o Y (vertikaloje) ašyje pateikiama PK3.

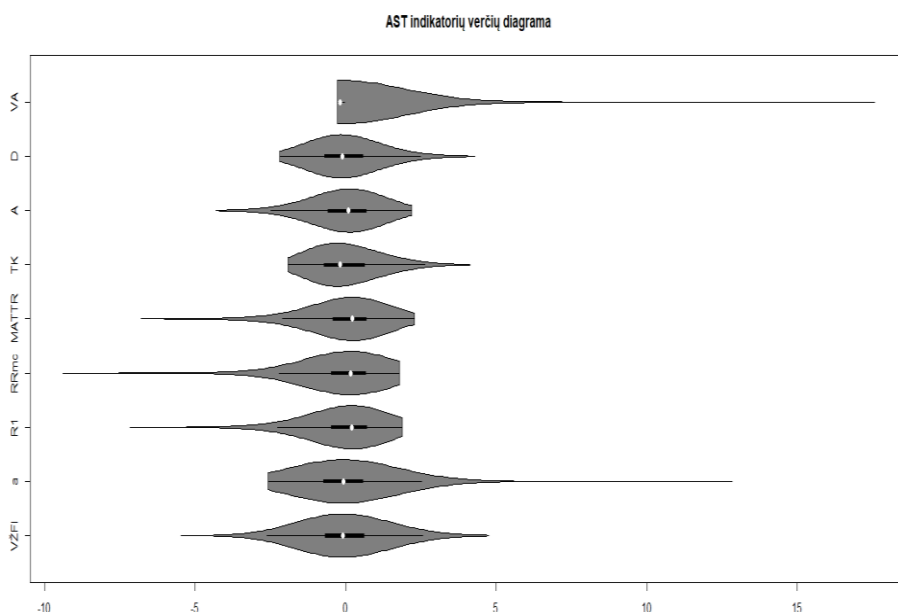


13 pav. PST: indikatorių indėlis: PK2 ir PK3

Taigi kaip matyti 13-ame paveiksle, bendrame komponentių PK2 ir PK3 kontekste indikatoriai R_1 , RR_{mc} ir TK daugiausiai prisideda apibūdinant PST – juos reprezentuojančios rodyklės yra raudoniausios, o jų smaigaliai baigiasi arčiausiai paveiksle pavaizduoto apskritimo perimetro. Indikatoriai R_1 , RR_{mc} apibūdina tekstą per žodžių ir žodžių formų įvairovės-koncentracijos dimensijas. Jų indėlis PST tekstų apibūdinimui aptartas anksčiau, analizuojant 12-ą paveikslą. Indikatorius TK įvertina, kaip tekstas yra sukoncentruotas į savo temą pagal dažnumu funkciniais žodžiams ir jų formoms beveik prilygstančių turinio žodžių ir jų formų skaičių. Analizuojant duomenis galima pastebėti, kad kitų indikatorių indėlis mažesnis (rodyklių smaigaliai arčiau apskritimo centro, o jų tolimesnės raudonos spalvos spektrui). Mažiausiai bendrame PK2 ir PK3 kontekste apibūdinant PST prisideda indikatorius VA, t. y. šių komponentių atžvilgiu indikatorius reprezentuoja tik labai mažą dalį PK2 ir PK3 duomenų reikšmių.

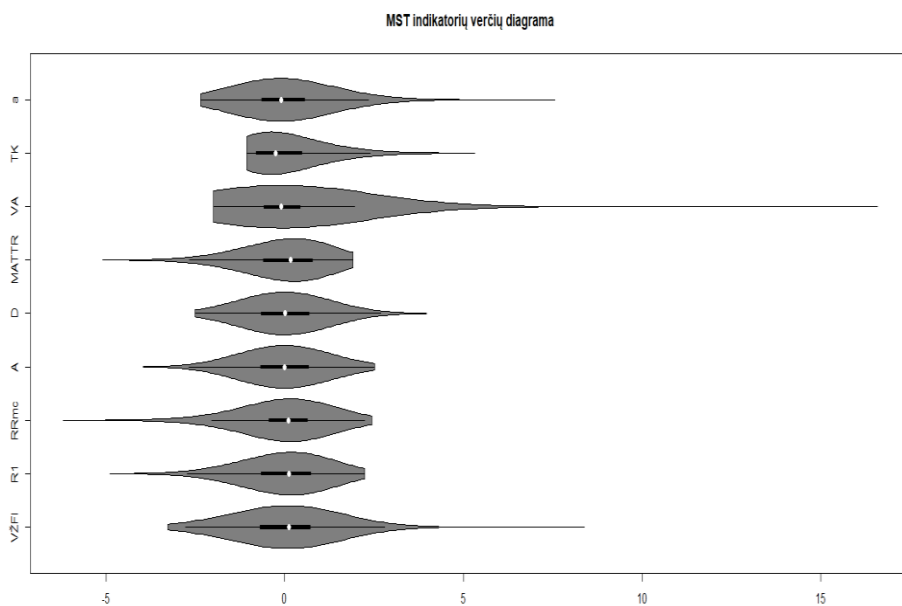
4.4 Atskirų FS kalbinės raiškos požymių analizės apibendrinimas¹¹¹

Šiame skyriuje apibendrinami atskirų FS kalbinės raiškos požymių tyrimo rezultatai. Siekiant susipažinti su tiriamais duomenimis – administraciniais tekstaais – ir jų savybėmis, buvo atlikta aprašomoji arba pirminė duomenų analizė. Aprašomosios duomenų analizės, šios metu analizuojamas duomenų verčių pasiskirstymas ir atliktos kiekvieno indikatoriaus atžvilgiu kiekvienam tiriamajam FS, apibendrinamieji rezultatai vizualiai pateikti 14-16 pav. smuiko (angl. *violin*) diagramose, o detaliau aprašyti 4.4.1-4.4.3 sk.

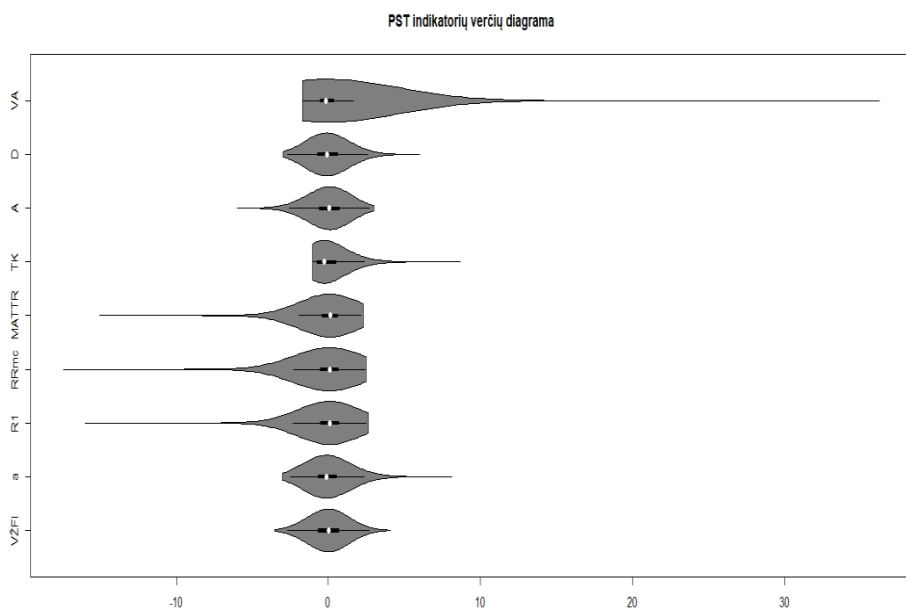


14 pav. Indikatorių verčių pasiskirstymo AST diagrama

¹¹¹ Šiame skyriuje apibendrinama atskirų FS kalbinės raiškos kiekybinė analizė, pačių stilių tarpusavyje nelyginant. FS kalbinės raiškos kiekybinė analizė, kuomet stiliai luginami tarpusavyje, pristatoma 5 sk.



15 pav. Indikatorių verčių pasiskirstymo MST diagrama



16 pav. Indikatorių verčių pasiskirstymo PST diagrama

Šiame skyriuje apibendrinami ir ryšių tarp indikatorių porų kiekvienam tiriamajam FS rezultatai. Taip pat pateikiami PKA rezultatai, leidę išskirti indikatorių, žyminčių kalbinės raiškos požymius, kombinacijas, geriausiai tinkančias apibūdinti atskirų FS kalbinę raišką.

4.4.1 Administracinio stiliaus kalbinės raiškos požymių kiekybinė analizės apibendrinimas

Vidutinis žodžio ir žodžio formos ilgis tekste. Administraciniams tekstams būdingesni vidutiniškai trumpesni žodžiai ir žodžių formos nei moksliniams tekstams, bet ilgesni nei publicistiniams tekstams. Kitaip tariant, vidutiniškai itin ilgi žodžiai ir žodžių formos šiuose tekstuose bus veikiau išimtis nei taisyklė. Kartu tai rodo, kad tekstų sudėtingumo (suprantamumo prasme) administraciniam stiliui tipiškesni suprantamumo prasme paprastesni tekstai nei moksliniam stiliui, bet sudėtingesni nei publicistiniam stiliui. Be to, pagal vidutinį žodžių ir žodžių formų ilgį administraciniai tekstai yra gana panašūs ir tarpusavyje pagal šį parametą skiriasi mažiau nei moksliniai ir publicistiniai tekstai.

Labai dažnų (paprastai – funkcinių) žodžių ir žodžių formų gausa tekste. Administracinio stiliaus tekstams būdingesnė šiek tiek didesnė labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių gausa nei publicistiniams tekstams, bet mažesnė nei moksliniams. Kita vertus, maža labai dažnų žodžių ir jų formų sritis administraciniuose tekstuose yra veikiau išimtis negu dėsningumas. Be to, šiems tekstams būdinga labai dažnų žodžių ir žodžių formų srities dažninėje žodžių ir žodžių formų lentelėje įvairovė, o tai rodo, kad pagal šį parametą administraciniai tekstai reikšmingai skiriasi tarpusavyje. Paminėtina, kad administraciniai tekstai pagal minėtą parametą skiriasi labiau nei publicistiniai, bet mažiau nei moksliniai.

Žodžių ir žodžių formų (retesnių) įvairovė tekste. Administraciniai tekstai, kuriems būdingesnė mažesnė (retesnių) žodžių ir jų formų įvairovė yra dažnesni nei tokie, kurių ši įvairovė didesnė. Kita vertus, itin maža (retesnių) žodžių ir jų formų įvairovė, tikėtina, bus išimtis, o ne tendencija. Be to, administracinis stilius iš dalies turėtų linkti link didesnės šio požymio raiškos įvairovės, nors, lyginant su kitais tiriamais funkciniais stiliais, pagal (retesnių) žodžių ir žodžių formų parametą administracinio stiliaus tekstai tarpusavyje skiriasi mažiau nei moksliniai ir publicistiniai.

Žodžių ir žodžių formų koncentruotumas tekste. Nors administraciniai tekstai, kuriems būdingesnis mažesnis žodžių ir žodžių formų koncentruotumas, yra dažnesni, tačiau labai mažas žodžių ir žodžių formų

koncentruotumas administracinio stiliaus tekstuose turėtų būti greičiau išimtis negu dėsningumas. Šio stiliaus tekstams būdinga maža žodžių ir žodžių formų koncentruotumo raiškos įvairovė, tad administraciniai tekstai šiuo atžvilgiu yra panašūs, nors vis tiek skirtingesni nei moksliniai ir publicistiniai.

Informacijos srautas (skirtingų žodžių ir žodžių formų kiekis tekste).

Administraciniams tekstams dažnesnis yra mažesnis nei vidurkis informacijos srautas, t. y. mažesnis nei vidurkis skirtingų žodžių ir žodžių formų kiekis, nors itin mažas bus greičiau išimtis, o ne tendencija. Be to, administracinio stiliaus tekstai informacijos srauto atžvilgiu tarpusavyje skiriasi vidutiniškai, tačiau labiau nei moksliniai ir publicistiniai.

Teksto teminė koncentracija. Itin didelė teminė koncentracija administraciniuose tekstuose turėtų būti greičiau išimtis nei dėsningumas, nors didesnė nei vidurkis teminė koncentracija šiuos tekstuose yra dažnesnė nei mažesnė. Šio požymio raiška administraciniuose tekstuose įvairuoja, t. y. pagal šį požymį šie tekstai yra tarpusavyje labai skirtingi ir šiuo atžvilgiu tarpusavyje skiriasi labiau nei moksliniai ir publicistiniai tekstai.

Teksto aktyvumas ir deskriptyvumas. Administraciniuose tekstuose dažnesni tekstai, kuriems būdingesnė mažesnė aktyvumo raiška (mažiau veiksmažodžių lyginant su būdvardžiais) ir didesnė deskriptyvumo raiška. Kita vertus, itin maža aktyvumo arba itin didelė deskriptyvumo raiška turėtų būti išimtiniai atvejai. Pagal šiuos parametrus administraciniai tekstai reikšmingai skiriasi tarpusavyje, ir tie tekstų tarpusavio skirtumai yra didesni nei moksliniams ir publicistiniams tekstams.

Teksto sintaksinės struktūros kompleksiskumas. Administraciniams tekstams itin didelis atstumas tarp 2 iš eilės einančių veiksmažodžių tekste, t. y. itin didelis teksto sintaksinės struktūros kompleksiskumas, nėra tipiškas, o greičiau išimtis. Kita vertus, didesni nei vidurkis vidutiniai atstumai tarp 2 iš eilės einančių veiksmažodžių yra dažnesni nei mažesni. Kitais žodžiais tariant, administraciniame stiliuje dažnesni yra sintaksinės struktūros požiūriu sudėtingesni tekstai. Be to, šiems tekstams būdinga sintaksės kompleksiskumo raiškos įvairovė, t. y. pagal šį požymį administraciniai tekstai ženkliai skiriasi tarpusavyje, nors ir mažiau nei moksliniai ir publicistiniai.

Ryšiai tarp požymių administraciniuose tekstuose. Administraciniuose tekstuose stipriausiai susiję yra aktyvumo ir deskriptyvumo bei (retesnių) žodžių ir žodžių formų įvairovės ir (visų) žodžių ir žodžių formų koncentruotumo požymiai. Labai stiprus ryšys tarp aktyvumo ir deskriptyvumo, kadangi šie požymiai yra vienas kitam priešingi, nustatytas ir publicistiniuose bei moksliniuose tekstuose. Ryšys tarp (retesnių) žodžių ir žodžių formų įvairovės ir (visų) žodžių ir žodžių formų koncentruotumo

moksliniuose tekstuose yra silpnesnis nei administraciniuose, o publicistiniuose – tokio pat stiprumo. Tad kuo administraciniame tekste didesnė aktyvumo raiška, tuo mažesnė – deskriptyvumo ir atvirkščiai. Tačiau kuo didesnė (retesnių) žodžių ir žodžių formų įvairovė, tuo didesnis ir (visų) žodžių ir žodžių formų koncentruotumas.

Be to, nustatyti vidutinio stiprumo ryšiai tarp kitų 8 požymių porų. Tad kuo administraciniame tekste mažiau (retesnių) žodžių ir žodžių formų, tuo jam būdingesnė didesnė labai dažnų žodžių gausa (ir atvirkščiai). Taip pat kuo šio stiliaus tekste didesnis žodžių ir jų formų koncentruotumas (kai vos keletas žodžių ir žodžių formų pasikartoja itin dažnai), tuo tekstui būdingesnė mažesnė labai dažnų žodžių gausa (ir atvirkščiai). Be to, kuo didesnis/mažesnis informacijos srautas, t. y. kuo daugiau tekste skirtingų žodžių ir jų formų, tuo mažesnė / didesnė labai dažnų žodžių ir žodžių formų gausa būdingesnė tekstui. Kita vertus, kuo tekstas pasižymi didesniu/mažesniu informacijos srautu (kitaip – didesniu skaičiumi skirtingų žodžių ir jų formų), tuo didesnė jame (retesnių) žodžių ir žodžių formų įvairovė. Kartu ir kuo didesnis/mažesnis informacijos srautas (tekste daugiau/mažiau skirtingų žodžių ir jų formų), tuo didesnis/mažesnis tekstą sudarančių žodžių ir jų formų koncentruotumas.

Taip pat kuo paprastesnė teksto sintaksinė struktūra, t. y. kuo mažesnis jame atstumas tarp 2 iš eilės einančių veiksmažodžių, tuo jam būdingesnis didesnis informacijos srautas ir atvirkščiai – kuo teksto sintaksinė struktūra sudėtingesnė, tuo administraciniame tekste mažesnis informacijos srautas (kitaip – tuo mažiau tekste skirtingų žodžių ir jų formų). Be to, kuo didesnė aktyvumo / dinamiškumo raiška būdingesnė tekstui (t. y. kuo daugiau jame veiksmažodžių lyginant su būdvardžiais), tuo paprastesnė jo sintaksinė struktūra (t. y. tuo vidutiniškai mažesnis tekste atstumas tarp 2 iš eilės einančių veiksmažodžių). Taip pat ir atvirkščiai: kuo mažesnė aktyvumo / dinamiškumo raiška administraciniame tekste, tuo jo sintaksinė struktūra yra sudėtingesnė. Galiausiai, kuo didesnė/mažesnė deskriptyvumo raiška, t. y. kuo daugiau/mažiau administracinio stiliaus tekste būdvardžių, lyginant su veiksmažodžiais, tuo teksto sintaksinė struktūra yra sudėtingesnė/paprastesnė, kitaip – tuo didesnis/mažesnis atstumas tekste tarp 2 iš eilės einančių veiksmažodžių.

Geriausiai administracinio stiliaus raišką apibūdinančių požymių kombinacija. Administracinio stiliaus tekstų kalbinę raišką geriausiai apibūdina ši požymių kombinacija (požymiai išrikiuoti svarbos tvarka pagal bendrą procentinį indėlį):

1. R_1 — (retesnių) žodžių ir žodžių formų įvairovė (iš viso 9,81 proc. (27,396 proc. PK1));
2. A — aktyvumas / dinamiškumas (9,53 proc (37,532 proc. PK2));
3. D — deskriptyvumas / aprašomumas (9,53 proc (37,532 proc. PK2));
4. RR_{mc} — žodžių ir žodžių formų koncentruotumas (iš viso 9,18 proc. (25,636 proc. PK1));
5. MATTR — informacijos srautas (iš viso 8,51 proc. (23,76 proc. PK1));
6. TK — teminė koncentracija (7,57 proc. (54,782 proc. PK3)).

Kitaip tariant, visi šie požymiai kartu geriausiai reprezentuoja administracinio stiliaus kalbinę raišką.

4.4.2 Mokslinio stiliaus kalbinės raiškos požymių kiekybinė analizės apibendrinimas

Vidutinis žodžių ir žodžių formų ilgis tekste. Moksliniams tekstams būdingesnis ilgesnis vidutinis žodžių ir žodžių formų ilgis nei administraciniams ir publicistiniams tekstams. Taip pat, nors moksliniams tekstams dažnesni vidutiniškai ilgesni žodžiai ir žodžių formos, šis polinkis administraciniams tekstams yra stipresnis. Be to, vidutiniškai itin trumpi žodžiai ir žodžių formos mokslinio stiliaus tekstams yra netipiški. Šių tekstų vidutinio žodžių ir žodžių formų ilgio raiška turėtų linkti didesnės įvairovės link. Kita vertus, pagal šį parametą, lyginant visus tiriamuosius funkcinis stilius, moksliniai tekstai tarpusavyje labiausiai, publicistiniai – mažiau nei moksliniai, o administraciniai tekstai šiuo atžvilgiu tarpusavyje skiriasi mažiausiai.

Labai dažnų (paprastai – funkcinių) žodžių ir žodžių formų gausa tekste. Mokslinio stiliaus tekstams būdingesnė didesnė labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių gausa nei administraciniams ir publicistiniams tekstams. Taip pat moksliniam stiliui dažnesnė mažesnė labai dažnų žodžių ir žodžių formų gausa, ir ši tendencija šiam stiliui yra stipresnė nei administraciniam ir publicistiniam. Be to, itin maža labai dažnų žodžių gausa moksliniams tekstams būtų netipiškas atvejis. Šie tekstai labai dažnų žodžių gausos raiškos atžvilgiu taip pat ženkliai skiriasi tarpusavyje. Lyginant su visais tiriamaisiais funkciniais stiliais, pagal šį parametą mokslinio stiliaus tekstai tarpusavyje skiriasi mažiau nei publicistiniai, bet labiau nei administraciniai.

Žodžių ir žodžių formų (retesnių) įvairovė tekste. Itin maža (retesnių) žodžių ir žodžių formų įvairovė moksliniams tekstams yra netipiška. Be to, šie tekstai žodžių ir žodžių formų įvairovės atžvilgiu tarpusavyje skiriasi iš

dalies, t. y. pagal šį parametą skirtumai tarp mokslinio stiliaus tekstų nėra labai ryškūs. Lyginant tiriamuosius tekstynus galima tarti, pagal (retesnių) žodžių ir žodžių formų įvairovę administraciniai tekstai tarpusavyje skiriasi labiau nei moksliniai, o publicistiniai – mažiau. Moksliniame stiliuje dažnesnė mažesnė (retesnių) žodžių ir žodžių formų įvairovė. Kita vertus, ši tendencija administraciniame ir publicistiniame stiliuose yra stipresnė nei moksliniame.

Žodžių ir žodžių formų koncentruotumas tekste. Didesnių ir mažesnių žodžių ir žodžių formų koncentruotumo atvejų moksliniuose tekstuose buvo maždaug po lygiai, tad juose neturėtų būti netipiškų raiškos atvejų (nebent keletas). Žodžių ir žodžių formų koncentruotumo atžvilgiu moksliniai tekstai tarpusavyje yra panašūs. Lyginant su kitais tiriamaisiais tekstynais galima tarti, kad jų visų 3 vidinė tekstų diferenciacija žodžių ir žodžių formų koncentruotumo atžvilgiu yra itin maža, nors administraciniai tekstai pagal šį požymį tarpusavyje skiriasi truputį labiau nei moksliniai, o publicistiniai – truputį mažiau nei moksliniai. Moksliniuose tekstuose dažnesnis mažesnis žodžių ir žodžių formų koncentruotumas. Šis mokslinių tekstų polinkis yra silpnesnis nei administracinių publicistinių, kitaip sakant, mažesnis žodžių ir žodžių formų koncentruotumas pastariesiems yra tipiškesnis nei moksliniams tekstams.

Informacijos srautas (skirtingų žodžių ir žodžių formų kiekis) tekste.

Itin mažas informacijos srautas mokslinio stiliaus tekstuose būtų išimtis, o ne tendencija. Be to, šio požymio raiška moksliniuose tekstuose yra artima vidurkiui, tad pagal šį požymį moksliniai tekstai tarpusavyje mažai skiriasi. Kita vertus, pagal šį požymį moksliniai tekstai tarpusavyje skiriasi labiau nei publicistiniai, bet mažiau nei administraciniai. Taip pat moksliniuose tekstuose dažnesnis mažesnis nei vidurkis informacijos srautas (mažesnis nei vidurkis skirtingų žodžių ir žodžių formų kiekis), nors mažesnis informacijos srautas administraciniams ir publicistiniams tekstams yra tipiškesnis nei moksliniams.

Teksto teminė koncentracija. Moksliniuose tekstuose itin didelė teminė koncentracija (didelis skaičius pagrindinems kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu nedaug nusileidžiančių funkciniams žodžiams ir žodžių formoms) nėra tipiška. Pagal šio požymio raišką moksliniai tekstai yra labai skirtingi. Paminėtina, kad visiems 3 tiriamiesiems tekstynams būdinga didelė vidinė diferenciacija teminės koncentracijos raiškos atžvilgiu, nors pagal šį požymį administraciniai tekstai tarpusavyje yra panašesni nei moksliniai, o publicistiniai tarpusavyje skiriasi labiau nei moksliniai. Be to, moksliniuose tekstuose dažnesnė didesnė teminė koncentracija, ir šis polinkis pastariesiems tekstams yra truputį stipresnis nei publicistiniams bei gerokai stipresnis nei administraciniams.

Teksto aktyvumas ir deskriptyvumas. Itin didelė aktyvumo / dinamiškumo (veiksmažodžių kiekis lyginant su būdvardžiais) bei deskriptyvumo / aprašomumo raiška moksliniams tekstams nėra tipiška. Be to, moksliniams tekstams būdinga aktyvumo ir deskriptyvumo raiškos įvairovė. Kitaip sakant, moksliniai tekstai pagal šiuos 2 parametrus tarpusavyje reikšmingai skiriasi. Kita vertus, pagal aktyvumo raišką moksliniai tekstai tarpusavyje skiriasi labiau pagal deskriptyvumo. Taip pat pagal aktyvumo ir deskriptyvumo raiškos įvairovę moksliniai tekstai tarpusavyje skiriasi labiau nei publicistiniai, bet mažiau nei administraciniai. Kartu didesnės ir mažesnės aktyvumo ir deskriptyvumo raiškos atvejų moksliniuose tekstuose yra daugmaž po lygiai. Tuo tarpu administraciniams ir publicistiniams tekstams būdingesnė mažesnė aktyvumo raiška nei moksliniuose tekstuose bei didesnė nei pastaruosiuose deskriptyvumo raiška.

Teksto sintaksinės struktūros kompleksškumas. Moksliniam stiliui dažnesni sintaksinės struktūros požiūriu kompleksiškesni nei vidurkis tekstai t. y. dažnesni tekstai, kuriuose vidutinis atstumas žodžių formomis tarp 2 iš eilės einančių veiksmažodžių yra didesnis nei vidurkis. Kita vertus, šiuo požiūriu itin sudėtingi moksliniai tekstai turėtų būti greičiau išimtis nei tipiški atvejai. Pagal šį požymį moksliniai tekstai tarpusavyje reikšmingai skiriasi. Be to, ir visų 3 tiriamųjų funkcinių stilių tekstai pagal sintaksinės struktūros kompleksškumą ženkliai skiriasi, nors mokslinio stiliaus tekstai, lyginant su kitais dviem, tarpusavyje skiriasi mažiausiai.

Ryšiai tarp požymių moksliniuose tekstuose. Moksliniuose tekstuose (kaip ir administraciniuose) stipriausiai susiję yra aktyvumo ir deskriptyvumo bei (retesnių) žodžių ir žodžių formų įvairovės ir (visų) žodžių ir žodžių formų koncentruotumo požymiai. Labai stiprus ryšys tarp aktyvumo ir deskriptyvumo, kadangi šie požymiai yra vienas kitam priešingi, nustatytas ir kitų dviejų tiriamųjų funkcinių stilių tekstuose. Žodžių ir žodžių formų (retesnių) įvairovės ir žodžių ir žodžių formų (visų) koncentruotumo moksliniuose tekstuose ryšys yra silpnesnis nei administraciniuose, o publicistiniuose – tokio pat stiprumo kaip administraciniuose. Tad kuo moksliniame tekste didesnė aktyvumo raiška (daugiau veiksmažodžių, lyginant su būdvardžiais), tuo mažesnė – deskriptyvumo (ir atvirkščiai). Bet kuo didesnė (retesnių) žodžių ir žodžių formų įvairovė, tuo didesnis ir (visų) žodžių ir žodžių formų koncentruotumas (ir atvirkščiai).

Taip pat nustatyti vidutinio stiprumo ryšiai tarp kitų 10 požymių porų. Taigi kuo moksliniam tekstui būdingesnis didesnis/mažesnis sudėtingumas suprantamumo prasme (arba kitaip – didesnis/mažesnis vidutinis žodžių ir žodžių formų ilgis), tuo šiam tekstui taip pat tipiškesnė didesnė/mažesnė (retesnių) žodžių ir žodžių formų įvairovė. Be to, kuo mokslinio stiliaus tekste

didesnė aktyvumo / dinamiškumo raiška (daugiau veiksmažodžių, lyginant su būdvardžiais), tuo jam būdingesni vidutiniškai trumpesni žodžiai ir žodžių formos (mažesnis sudėtingumas suprantamumo prasme). Kita vertus, kuo didesnė/mažesnė deskriptyvumo raiška, tuo didesnis ir teksto sudėtingumas (būdingesni vidutiniškai ilgesni žodžiai ir žodžių formos). Kartu ir kuo didesnė (retesnių – priklausančių pagrindinėms kalbos dalims) žodžių ir žodžių formų įvairovė tekste, tuo jam būdingesnė mažesnė labai dažnų žodžių gausa (ir atvirkščiai).

Taip pat kuo mokslinių tekstų sintaksinė struktūra kompleksiškesnė (kitais tariant, kuo šiuose tekstuose atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis), tuo mažesnė juose aktyvumo raiška, t. y. tuo mažiau juose, lyginant su būdvardžiais, yra veiksmažodžių bei atvirkščiai. Be to, kuo daugiau tekste veiksmažodžių, lyginant su būdvardžiais, tuo jame mažiau retesnių, pagrindinėms kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu bemaž prilygstančių funkciniams. Kitais tariant, kuo didesnė tekste aktyvumo / dinamiškumo raiška, tuo mažesnė jo teminė koncentracija. Kita vertus, kuo teksto sintaksinė struktūra kompleksiškesnė/paprastesnė (kitais tariant, kuo tekste atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis/mažesnis), tuo jame deskriptyvumo raiška didesnė/mažesnė (t. y. tuo daugiau/mažiau tekste būdvardžių, lyginant su veiksmažodžiais). Tačiau kuo tekste daugiau būdvardžių, lyginant su veiksmažodžiais, tuo jame daugiau funkciniams žodžiams dažnumu nedaug nusileidžiančių retesnių, pagrindinėms kalbos dalims priklausančių, žodžių ir žodžių formų. Kitais žodžiais tariant, kuo didesnis teksto deskriptyvumas, tuo tam būdingesnė ir didesnė teminė koncentracija. Kartu ir kuo didesnis informacijos srautas (t. y. kuo didesnis skaičius skirtingų žodžių ir žodžių formų) būdingas tekstui, tuo mažesnė jo teminė koncentracija (kitais – tuo mažiau jame iš esmės retesnių turinio/teminių, tačiau šiame tekste savo dažnumu funkciniams bemaž prilygstančių, žodžių ir žodžių formų) ir atvirkščiai. Galiausiai, kuo didesnė/mažesnė teksto teminė koncentracija, tuo didesnis/mažesnis teksto sudėtingumas suprantamumo prasme, t. y. tuo ilgesnis/trumpesnis tekste vidutinis žodžių ir jų formų ilgis.

Geriausios mokslinio stiliaus raišką apibūdinančių požymių kombinacija. Mokslinio stiliaus tekstų kalbinę raišką geriausiai apibūdina ši požymių kombinacija (požymiai išrikiuoti svarbos tvarka pagal bendrą procentinę indėlį):

1. A — aktyvumas / dinamiškumas (8,79 proc. (24,137 proc. PK1));
2. D — deskriptyvumas / aprašomumas (8,79 proc. (24,137 proc. PK1));

3. RR_{mc} — žodžių ir žodžių formų koncentruotumas (7,79 proc. (31,519 proc. PK2));
4. R_1 — (retesnių) žodžių ir žodžių formų įvairovė (6,84 proc. (27,682 proc. PK2));
5. VA — teksto sintaksinės struktūros kompleksiskumas (3,6 proc. (27,997 proc. PK3)).

Kitaip sakant, visi šie požymiai kartu geriausiai reprezentuoja mokslinio stiliaus kalbinę raišką.

4.4.3 Publicistinio stiliaus kalbinės raiškos požymių kiekybinė analizė apibendrinimas

Vidutinis žodžių ir žodžių formų ilgis tekste. Publicistiniams tekstams būdingesni vidutiniškai trumpesni žodžiai ir žodžių formos nei administraciniais ir moksliniams tekstams. Be to, publicistiniams tekstams netipiški vidutinio žodžių ir žodžių formų ilgio atvejai (kai vidutinis žodžių ir žodžių formų ilgis yra itin didelis arba mažas) nėra būdingi. Pagal vidutinį žodžių ar žodžių formų ilgį šie tekstai iš dalies skiriasi tarpusavyje. Kita vertus, pagal šį parametą publicistiniai tekstai tarpusavyje skiriasi labiau nei administraciniai, bet mažiau nei moksliniai. Publicistiniams tekstams šiek tiek dažnesni, nors tas polinkis ir nedidelis, vidutiniškai trumpesni nei vidurkis žodžiai ir žodžių formos.

Labai dažnų (paprastai – funkcinų) žodžių ir žodžių formų gausa tekste. Publicistinio stiliaus tekstams būdingesnė mažesnė labai dažnų (paprastai – funkcinų, bet nebūtinai) žodžių gausa nei administraciniais ir moksliniams tekstams. Kita vertus, itin maža labai dažnų žodžių ir jų formų gausa publicistiniams tekstams nėra tipiška, nors mažesnė jos raiška dažnesnė. Šiems tekstams taip pat būdinga labai dažnų žodžių ir jų formų gausos raiškos įvairovė, tad publicistiniai tekstai tarpusavyje yra labai skirtingi. Publicistinio stiliaus tekstai tarpusavyje skiriasi labiau nei administraciniai ir moksliniai.

Žodžių ir žodžių formų (retesnių) įvairovė tekste. Publicistiniuose tekstuose dažnesnė mažesnė (retesnių) žodžių ir jų formų įvairovė, nors itin didelė jos raiška publicistinio stiliaus tekstuose, tikėtina, bus tik pavieniai atvejai, o ne tendencija. Kita vertus, šiuo tekstu kalbinė raiška (retesnių) žodžių ir žodžių formų įvairovės kontekste yra panaši, jie nelabai skiriasi vienas nuo kito. Taip pat publicistiniai tekstai pagal šį parametą tarpusavyje skiriasi mažiau nei administraciniai ir publicistiniai.

Žodžių ir žodžių formų koncentruotumas tekste. Publicistiniuose tekstuose dažnesnė mažesnis žodžių ir jų formų koncentruotumas, tačiau itin mažos jo raiškos atvejai turėtų būti netipiškos išimtis. Pagal šio požymio raiškos įvairovę publicistiniai tekstai tarpusavyje mažai skiriasi (raiška artima vidurkiui). Be to, pagal žodžių ir žodžių formų koncentruotumą publicistiniai tekstai tarpusavyje skiriasi mažiau nei administraciniai ir moksliniai.

Informacijos srautas (skirtingų žodžių ir žodžių formų kiekis) tekste. Publicistinio stiliaus tekstuose dažnesnis mažesnis nei vidurkis informacijos srautas, kitaip – mažesnis nei vidurkis skirtingų žodžių ir žodžių formų skaičius. Kita vertus, itin didelis informacijos srautas (itin didelis kiekis skirtingų žodžių ir jų formų) publicistiniuose tekstuose nėra tipiškas. Šio požymio raiškos įvairovė maža, todėl publicistiniai tekstai pagal jį nedaug skiriasi tarpusavyje. Taip pat pagal informacijos srauto (skirtingų žodžių ir jų formų kiekį tekste) publicistiniai tekstai tarpusavyje skiriasi mažiau nei administraciniai ir moksliniai.

Teksto teminė koncentracija. Publicistiniuose tekstuose teminė koncentracija dažniau yra didesnė, tačiau itin temiškai koncentruoti (t. y. tokie tekstai, kuriems būdinga pagrindinėms kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu nedaug nusileidžiančių funkciniam žodžiams ir žodžių formoms, gausa) publicistiniame stiliuje turėtų būti išimtis, o ne vyraujanti tendencija. Šio požymio raiškai būdinga įvairovė, tad publicistiniai tekstai pagal teminę koncentraciją labai skiriasi tarpusavyje. Kita vertus, pagal šį požymį publicistiniai tekstai tarpusavyje skiriasi labiau nei moksliniai, bet mažiau nei administraciniai.

Teksto aktyvumas ir deskriptyvumas. Publicistiniame stiliuje dažnesni tekstai, kurių aktyvumo raiška mažesnė (mažiaus veiksmažodžių lyginant su būdvardžiais), o deskriptyvumo – didesnė (daugiau būdvardžių lyginant su veiksmažodžiais). Kita vertus, itin mažai dinamiški bei itin didelės deskriptyvumo raiškos publicistiniai tekstai neturėtų vyrauti, o būtų tik išimtis. Be to, publicistiniams tekstams būdinga abiejų požymių atžvilgiu raiškos įvairovė, tad šie tekstai pagal aktyvumą (veiksmažodžių kiekis lyginant su būdvardžiais) ir deskriptyvumą gana reikšmingai skiriasi tarpusavyje. Taip pat pagal šiuos 2 požymius publicistiniai tekstai tarpusavyje skiriasi mažiau nei administraciniai ir moksliniai.

Teksto sintaksinės struktūros kompleksškumas. Publicistiniame stiliuje dažnesni yra sintaksės požiūriu paprastesni tekstai, t. y. tokie, kurių vidutinis atstumas tarp 2 iš eilės einančių veiksmažodžių yra mažesnis nei vidurkis. Kita vertus, labai sudėtingos sintaksinės struktūros tekstai (t. y. tokie, kuriose vidutinis atstumas tarp dviejų iš eilės einančių veiksmažodžių bus labai didelis), tikėtina, priklausys išimtimis. Publicistiniams tekstams būdinga

sintaksinės struktūros kompleksiskumo raiškos įvairovė, todėl šie tekstai gana ženkliai skiriasi tarpusavyje. Be to, publicistiniai tekstai sintaksinės struktūros kompleksiskumo kontekste tarpusavyje skiriasi šiek tiek mažiau nei moksliniai ir gerokai mažiau nei administraciniai.

Ryšiai tarp požymių publicistiniuose tekstuose. Publicistiniuose tekstuose (kaip ir administraciniuose bei moksliniuose) stipriausiai susiję yra aktyvumo ir deskriptyvumo bei (retesnių) žodžių ir žodžių formų įvairovės ir (visų) žodžių ir žodžių formų koncentruotumo požymiai. Labai stiprus ryšys tarp aktyvumo ir deskriptyvumo, kadangi šie požymiai yra vienas kitam priešingi, nustatytas visų tiriamųjų funkcinių stilių tekstuose. Žodžių ir žodžių formų (retesnių) įvairovės ir žodžių ir žodžių formų (visų) koncentruotumo ryšys publicistiniuose ir administraciniuose tekstuose yra tokio pat stiprumo, moksliniuose – silpnesnis. Taigi kuo publicistiniame tekste didesnė aktyvumo raiška (daugiau veiksmažodžių, lyginant su būdvardžiais), tuo mažesnė – deskriptyvumo (ir atvirkščiai). Tačiau kuo didesnė/mažesnė (retesnių) žodžių ir žodžių formų įvairovė, tuo didesnis/mažesnis ir (visų) žodžių ir žodžių formų koncentruotumas.

Nustatyti ir vidutinio stiprumo ryšiai tarp kitų 11 požymių porų. Tad didėjant/mažėjant teminei koncentracijai (didėjant/mažėjant tekste pagrindinėms kalbos dalims priklausančių žodžių ir žodžių formų, dažnumu nedaug nusileidžiančių funkciniams žodžiams ir žodžių formoms), teksto sudėtingumas taip pat didėja/mažėja, t. y. ilgėja/trumpėja vidutinis žodžių ir jų formų ilgis. Taip pat didėjant teksto aktyvumo (kitaip – daugėjant tekste veiksmažodžių, lyginant su būdvardžiais), jo sudėtingumas mažėja (kitaip – vidutinis žodžių ir jų formų ilgis trumpėja) ir atvirkščiai. Tačiau didėjant/mažėjant teksto deskriptyvumo raiškai (t. y. daugėjant/mažėjant tekste būdvardžių, lyginant su veiksmažodžiais), jo sudėtingumas taip pat didėja/mažėja (t. y. ilgėja/trumpėja vidutinis žodžių ir žodžių formų ilgis). Be to, kuo sudėtingesnė/paprastesnė teksto sintaksinė struktūra (t. y. didesnis/mažesnis atstumas tarp 2 iš eilės einančių veiksmažodžių), tuo šis tekstas yra sudėtingesnis (sunkiau suprantamas/ paprastesnis (lengviau suprantamas), t. y. tuo jam būdingesni vidutiniškai ilgesni/trumpesni žodžiai ir žodžių formos.

Kartu kuo didesnė/mažesnė tiriamų publicistinių tekstų (retesnių) žodžių ir jų formų įvairovė, tuo jiems būdingesnė mažesnė / didesnė labai dažnų (paprastai – funkcinių) žodžių ir žodžių formų gausa. Taip pat kuo didesnis/mažesnis žodžių ir jų formų koncentruotumas tekste, tuo jam būdingesnė mažesnė/didesnė labai dažnų žodžių gausa. Be to, kuo didesnis/mažesnis informacijos srautas tekste (kitaip – kuo daugiau / mažiau tekste skirtingų žodžių ir žodžių formų), tuo tekstui būdingesnė mažesnė /

didesnė gausa labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų. Lygiai taip pat, kuo didesnis/mažesnis informacijos srautas tekste (kitai sakant, kuo daugiau tekste skirtingų žodžių ir žodžių formų), tuo jo (retesnių) žodžių ir žodžių formų įvairovė yra didesnė/mažesnė. Kartu kuo didesnis/mažesnis informacijos srautas būdingas tekstui, tuo tipiškesnis jam didesnis/mažesnis žodžių ir žodžių formų koncentruotumas. Kita vertus, kuo publicistinių tekstų sintaksinė struktūra kompleksiškesnė (kitai tariant, kuo šiuose tekstuose atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis), tuo mažesnė juose aktyvumo raiška, t. y. tuo mažiau juose, lyginant su būdvardžiais, juose yra veiksmažodžių bei atvirkščiai. Galiausiai, kuo teksto sintaksinė struktūra kompleksiškesnė/paprastesnė (kitai tariant, kuo tekste atstumas tarp 2 iš eilės einančių veiksmažodžių didesnis/mažesnis), tuo jame deskriptyvumo raiška didesnė/mažesnė (t. y. tuo daugiau/mažiau tekste būdvardžių, lyginant su veiksmažodžiais).

Geriausiai publicistinio stiliaus raišką apibūdinančių požymių kombinacija. Publicistinio stiliaus tekstų kalbinę raišką geriausiai apibūdina ši požymių kombinacija (požymiai išrikiuoti svarbos tvarka pagal bendrą procentinį indėlį):

1. R_1 — (retesnių) žodžių ir žodžių formų įvairovė (7,15 proc. (22,612 proc. PK2));
2. VA — teksto sintaksinės struktūros kompleksiskumas (6,72 proc. (18,934 proc. PK1));
3. A — aktyvumas / dinamiškumas (6,59 proc. (18,574 proc. PK1));
4. D — deskriptyvumas / aprašomumas (6,59 proc. (18,574 proc. PK1));
5. TK — teminė koncentracija (6,37 proc. (51,281 proc. PK3));
6. RR_{mc} — žodžių ir žodžių formų koncentruotumas (5,98 proc. (18,920 proc. PK2)).

Kitai tariant, visi šie požymiai kartu geriausiai reprezentuoja publicistinio stiliaus kalbinę raišką.

5. TIRIAMŪJŲ FS PALYGINIMAS

5.1 FS tarpusavio panašumų ir skirtumų pagal kalbinę raišką žyminčius indikatorius analizė

Atlikus PKA, per kiekybinius indikatorius įvertinti atskirų tiriamų FS būdingesni požymiai ir jų kombinacijos. FS – kalbos kaip sistemos dalis, todėl šiame skyriuje aprašomame tyrimo etape tirti administracinio, mokslinio ir publicistinio FS tarpusavio panašumai ir skirtumai. Tam pasirinkti neparametriniai išvadų statistikos testai, leidę nustatyti tiek bendrą tiriamų FS tarpusavio skirtumų atskirų kiekybinių indikatorių atžvilgiu laipsnį, tiek įvertinti šiuos skirtumus stilių porose bei išskirti būdingesnes indikatorių verčių, o kartu ir juos atitinkančių FS skiriamųjų požymių, raiškos tendencijas.

30 lentelė. Kruskal'o-Wallis'o testo rezultatai

Indikatorius	Chi kvadrato vertė	<i>p</i> -vertė
VŽFI	10051	< 2.2e-16
<i>a</i>	9127,3	< 2.2e-16
R ₁	9883,2	< 2.2e-16
RR _{mc}	9307	< 2.2e-16
MATTR	10572	< 2.2e-16
TK	5496,2	< 2.2e-16
A	1656,9	< 2.2e-16
D	1656,9	< 2.2e-16
VA	6139	< 2.2e-16

Visų pirma, siekta nustatyti, ar tiriami FS apskritai skiriasi tarpusavyje. Kitaip sakant, norėta užfiksuoti skirtumų tarp 3 stilių buvimą arba nebuvimą pagal atskirus kiekybinius indikatorius. Tam naudotas Kruskal'o-Wallis'o testas (Field, Miles ir Field, 2012, 674—677; Mangiafico, 2016, 248—249), kurio rezultatai pateikti 30-oje lentelėje. Šio testo rezultatai parodė, kad skirtumai yra statistiškai reikšmingi visiems tiriamiesiems FS ir pagal visus tyrime naudotus indikatorius (visos *p*-vertės yra mažesnės nei 0,05). Šių skirtumų įvertinimui naudoti kiti 2 testai – *epsilon* kvadratas (Tomczak ir Tomczak, 2014) bei VDA (Vargha ir Delaney, 2000).

31 lentelė. *Epsilon* kvadrato testo rezultatai

Indikatorius	<i>Epsilon</i> kvadrato vertė
VŽFI	0,526
<i>a</i>	0,478
R ₁	0,517
RR _{mc}	0,487
MATTR	0,553
TK	0,288
A	0,0867
D	0,0867
VA	0,321

31-oje lentelėje pateikti *epsilon* kvadrato testo rezultatai. Kuo *epsilon* kvadrato vertė tam tikram indikatoriumi didesnė, tuo tiriami FS pagal jį labiau skiriasi tarpusavyje. Tiriami FS tarpusavyje labiausiai skiriasi pagal MATTR¹¹² (0,553), VŽFI¹¹³ (0,526) ir R₁¹¹⁴ (0,517). Mažiausiai tiriami FS tarpusavyje skiriasi pagal A¹¹⁵ ir D¹¹⁶ vertes (0,0867). Kitaip tariant, didžiausi skirtumai tarp stilių randami informacijos srauto, teksto suprantamumo sudėtingumo bei žodžių ir žodžių formų įvairovės atžvilgiais. Mažiausi – teksto „tono“ atžvilgiu (aktyvumo ir deskriptyvumo dimensijos). Kad būtų galima įvertinti, kurie tiriami funkciniai stiliai pagal kurį indikatoriumi skiriasi labiausiai, toliau taikytas VDA testas (Vargha ir Delaney, 2000).

32 lentelė. VDA testo rezultatai

Indikatorius	VDA statistika	Labiausiai besiskiriančių FS tekstynų pora
VŽFI	0,98	AST-PST
<i>a</i>	0,9473	AST-PST
R ₁	0,9822	MST-PST
RR _{mc}	0,9674	MST-PST
MATTR	0,9852	AST-PST

¹¹² Informacijos srauto įvertis (arba vidutinio skirtingų žodžių ir žodžių formų kiekio tekste matas), plačiau žr. 1.4.2.4 sk.

¹¹³ Vidutinis žodžių ar žodžių formų ilgis tekste (arba teksto sudėtingumo suprantamumo prasme įvertis), plačiau žr. 1.4.3.1 sk.

¹¹⁴ Žodžių ir žodžių formų (retesnių) įvairovės indikatorius, plačiau žr. 1.4.2.2 sk.

¹¹⁵ Aktyvumo indikatorius (veiksmazodžių ir būdvardžių santykis tekste), plačiau žr. 1.4.3.2 sk.

¹¹⁶ Deskriptyvumo indikatorius, plačiau žr. 1.4.3.2 sk.

TK	0,944	AST-MST
A	0,835	MST-PST
D	0,835	MST-PST
VA	0,876	AST-PST

32-ioje lentelėje galima matyti, kurios tiriamų FS poros atskirų indikatorių atžvilgiu skiriasi labiausiai. Pagal tekstų suprantamumą / sudėtingumą (indikatorius VŽFI), labai dažnų žodžių ir jų formų dalies tekste didumą (indikatorius a), informacijos srauto dydį (indikatorius MATTR) ir tekstų sintaksinį kompleksumą (indikatorius VA) labiausiai skiriasi AST ir PST. Žodžių ir žodžių formų įvairovė (indikatorius R_1) bei koncentracija (indikatorius RR_{mc}) ir aktyvumo bei deskriptyvumo raiška labiausiai skiria MST ir PST. Galiausiai, pagal teminės koncentracijos raišką patys skirtingiausi yra AST ir MST.

33 lentelė. Dunn'o testo rezultatai

Indikatorius	Lyginama FS tekstynų pora	Z vertė	<i>p</i> -vertė (adaptuota daugybiniam palyginimams)
VŽFI	AST-MST	18.36141	8.027132e-75
	AST-PST	98.41522	0.000000e+00
	MST-PST	32.20224	4.925667e-227
a	AST-MST	-14.32894	4.329066e-46
	AST-PST	-93.11607	0.000000e+00
	MST-PST	-33.71111	1.191995e-248
R_1	AST-MST	3.496798	0.001412636
	AST-PST	-91.122320	0.000000000
	MST-PST	-51.653576	0.000000000
RR_{mc}	AST-MST	0.350791	1
	AST-PST	-89.598949	0
	MST-PST	-47.500637	0
MATTR	AST-MST	-19.85568	2.952350e-87
	AST-PST	-101.12481	0.000000e+00
	MST-PST	-32.03546	1.050061e-224
TK	AST-MST	45.73392	0.000000e+00
	AST-PST	71.06957	0.000000e+00
	MST-PST	-11.34295	2.411528e-29
A	AST-MST	18.98684	6.574064e-80

	AST-PST	-26.51339	2.038068e-154
	MST-PST	-34.17354	1.793499e-255
D	AST-MST	18.98684	6.574064e-80
	AST-PST	-26.51339	2.038068e-154
	MST-PST	-34.17354	1.793499e-255
VA	AST-MST	17.86002	7.246323e-71
	AST-PST	77.51638	0.000000e+00
	MST-PST	21.74412	2.355814e-104

Jei Kruskal'o-Wallis'o testas leido nustatyti tik skirtumų tarp tiriamų FS tekstynų egzistavimą pagal atskirus indikatorius, o VDA – įvertinti labiausiai tarpusavyje besiskiriančias stilių poras, tai Dunn'o testo tikslas – rasti, ar egzistuoja statistiškai reikšmingi skirtumai tarp visų galimų FS tekstynų porų. Šio testo rezultatai pateikti 33-oje lentelėje. Visos galimos FS tekstynų poros (AST-MST, AST-PST ir MST-PST) tarpusavyje skiriasi statistiškai reikšmingai pagal visus tyrime naudojamus indikatorius, išskyrus AST-MST porą. Žodžių ir žodžių formų koncentracijos (indikatorius RR_{mc}) atžvilgiu tarp AST ir MST esantys skirtumai yra statistiškai nereikšmingi (adaptuota daugybiniais palyginimams p -vertė lygi 1). Tai rodo, kad žodžių ir žodžių formų koncentracijos raiška administraciniuose ir moksliniuose tekstuose yra panaši. Kitaip tariant, šiuose tekstuose žodžių ir žodžių formų pasikartojimų modelis yra mažai skiriasi.

34 lentelė. Santykiniai poveikio įverčiai

Indikatorius	Tekstynas	Santykiniai poveikio įverčiai
VŽFI	AST	0,85076
	MST	0,66777
	PST	0,36655
<i>a</i>	AST	0,16975
	MST	0,31255
	PST	0,62789
R_1	AST	0,18619
	MST	0,15135
	PST	0,63452
RR_{mc}	AST	0,18979
	MST	0,18629
	PST	0,63062
MATTR	AST	0,13905
	MST	0,33693

	PST	0,63660
TK	AST	0,77057
	MST	0,31495
	PST	0,42102
A	AST	0,41830
	MST	0,22908
	PST	0,54875
VA	AST	0,77809
	MST	0,60010
	PST	0,39670

Toliau aptarsime indikatorių verčių tendencijas tiriamiems FS. Šių tendencijų skaitinės išraiškos pateiktos 34-oje lentelėje. Čia pateikti santykiniai poveikio įverčiai rodo, kokia yra tikimybė, kad tekstui, priklausančiam tam tikram FS, bus būdingesnės didesnės konkretaus indikatoriaus vertės lyginant su visais tiriamaisiais tekstais. Kuo šie santykiniai poveikio įverčiai didesni, tuo minėtoji tikimybė yra didesnė. O kuo santykinis poveikio įverčių reikšmės mažesnės, tuo ta tikimybė yra mažesnė.

Kaip rodo rezultatai, administraciniams tekstams būdingesnės didesnės VŽFI vertės (suprantamumo požiūriu sudėtingesni tekstai). Kitaip sakant, tikimybė, kad atsitiktinis tekstas, priklausantis AST tekstynui, turės didesnę VŽFI vertę nei atsitiktinis tekstas iš visos tiriamųjų tekstų aibės, yra, 0,85076 arba 85,076 %. Šiuo atžvilgiu publicistiniams tekstams, priešingai, būdingesni suprantamumo požiūriu paprastesni tekstai nei administraciniams ir moksliniams (santykiniai poveikio įverčiai yra atitinkamai 0,36655, 0,85076 ir 0,66777). Moksliniai tekstai šiuo požiūriu yra per vidurį tarp administracinių ir publicistinių. Įvertinant tikimybes galima sakyti, jog pagal VŽFI administraciniai tekstai yra labiausiai nutolę nuo publicistinių tekstų. Kita vertus, šiuo atžvilgiu moksliniai tekstai yra artimesni administraciniams nei publicistiniams.

Publicistiniams tekstams būdingesnės didesnės indikatoriaus *a* vertės, t. y. mažesnė labai dažnų žodžių ir jų formų sritis, o administraciniams, priešingai, tipiškesnė didesnė labai dažnų žodžių sritis (santykiniai poveikio įverčiai yra atitinkamai 0,62789 ir 0,16975). Moksliniai tekstai šiuo požiūriu yra tarp administracinio ir publicistinio (santykiniai poveikio įvertis yra 0,31255). Tad labai dažnų žodžių ir jų formų srities dydžio atžvilgiu publicistiniai tekstai yra labiausiai nutolę nuo administracinių, o moksliniai tekstai pagal šį požymį yra artimesni administraciniams tekstams nei publicistiniams.

Pagal žodžių ir jų formų įvairovės požymį (indikatorius R_1) publicistiniams tekstams būdingesnė didesnė šio požymio raiška (didesnės R_1 vertės) nei kitiems 2 tiriamiems funkciniais stiliams (santykinio poveikio įverčio reikšmė – 0,63452). Moksliniams tekstams būdingesnė mažesnė žodžių ir jų formų įvairovė (mažesnės R_1 vertės; santykinio poveikio įverčio reikšmė – 0,15135). Kita vertus, nors pagal šį požymį publicistiniai tekstai yra labiausiai nutolę nuo mokslinių, tačiau skirtumas tarp mokslinių ir administracinių tekstų yra nedidelis. 34-oje lentelėje galima matyti, jog MST tekstams R_1 santykinio poveikio įverčio reikšmė, kaip minėta, yra 0,15135, o AST tekstams – 0,18619. Vis tik skirtumas tarp MST ir AST nors nedidelis, bet statistiškai reikšmingas (žr. Dunn'o testo rezultatus 33-oje lentelėje).

Kaip rodo rezultatai, pagal RR_{mc} – žymi žodžių ir žodžių formų koncentraciją tekste – didesnės vertės būdingesnės publicistiniams tekstams (santykinio poveikio įverčio reikšmė – 0,63062), o mažesnės – administraciniams ir moksliniams tekstams (santykinio poveikio įverčio reikšmė atitinkamai 0,18979 ir 0,18629). Kadangi šių stilių santykinio poveikio įverčio reikšmės labai nedaug skiriasi, galima sakyti, kad pagal šį požymį administraciniai ir moksliniai tekstai yra panašūs. Dunn'o testo rezultatai irgi parodė, kad RR_{mc} atžvilgiu skirtumas tarp šių 2 FS yra statistiškai nereikšmingas (žr. 33-ą lentelę).

Publicistiniams tekstams būdingesnis ir didesnis informacijos srautas (arba vidutiniškai didesnis skaičius skirtingų žodžių), t. y. didesnės indikatoriaus MATTR vertės (santykinio poveikio įverčio reikšmė – 0,63660). Mažesnė šio požymio raiška būdingesnė administraciniams tekstams (santykinio poveikio įverčio reikšmė – 0,13905). Moksliniai tekstai šiuo atžvilgiu yra tarp administracinių ir publicistinių, nors ir šiek tiek arčiau pirmųjų. Kitaip sakant, skirtumas tarp MST ir AST santykinio poveikio įverčio reikšmių (atitinkamai 0,33693 ir 0,13905) yra mažesnis nei tarp MST ir PST (atitinkamai 0,33693 ir 0,63660).

Administraciniams tekstams būdingesnė didesnė teminė koncentracija nei kitiems tiriamiems FS (santykinio poveikio įverčio reikšmė – 0,77057). Kitaip sakant, administraciniams tekstams būdingesni teminiai / turinio žodžiai ir žodžių formos¹¹⁷, savo dažnumu nedaug nusileidžiantys funkciniais žodžiams, kurie paprastai tekste būna dažniausi. Mažesnė teminė koncentracija arba mažesnis kiekis teminių / turinio žodžių ir žodžių formų, savo dažnumu bemaž prilygstančių funkciniais žodžiams, būdingesnis

¹¹⁷ Tai žodžiai ir žodžių formos, priklausantys pagrindinėms kalbos dalims (plačiau žr. 1.4.2.5 sk.).

moksliniams tekstams (santykinio poveikio įverčio reikšmė – 0,39670). Teminės koncentracijos atžvilgiu tiriamieji publicistiniai tekstai yra artimesni moksliniams tekstams nei administraciniais (santykinio poveikio įverčio reikšmė – 0,42102).

Kaip buvo minėta indikatorių ryšių analizės atskiriems tiriamiems FS skyriuose (žr. 4.1.2, 4.2.2 ir 4.3.2 skyrius), indikatorius A ir D sieja labai stiprus ryšys (Spirmeno koreliacijos koeficiento vertė -1). Kai priklausomi kintamieji yra itin stipriai koreliuoti, daugelis analizės metodų (dispersinės, faktorinės, regresinės ir panašiai) negali įvertinti jų indėlio apibūdinant nepriklausomą kintamąjį (Field, Miles ir Field, 2012, 770—771). Disertacijoje atliekamo tyrimo kontekste tai reiškia, kad, kadangi indikatoriai A ir D yra labai stipriai tarpusavyje susiję, atliekant tiriamų FS tarpusavio panašumų ir skirtumų analizę nėra galimybės įvertinti šių dviejų indikatorių indėlio charakterizuojant tiriamus stilius. Paprastai tokio pobūdžio problema sprendžiama pašalinant vieną kintamąjį iš labai stipriai koreliuotų kintamųjų poros (Field, Miles ir Field, 2012, 771—772). Tad santykinio poveikio įverčių skaičiavimui iš indikatorių poros A ir D pasiliktas indikatorius A, vertinantis teksto aktyvumą. Pašalintasis indikatorius D, vertinantis teksto deskriptyvumą, yra priešingas indikatoriumi A. Kitaip sakant, jei tekstui būdinga didelė aktyvumo raiška, tai tuo pačiu jam būdinga ir maža deskriptyvumo raiška, tad atsisakius indikatoriaus D iš esmės nedaug prarandama.

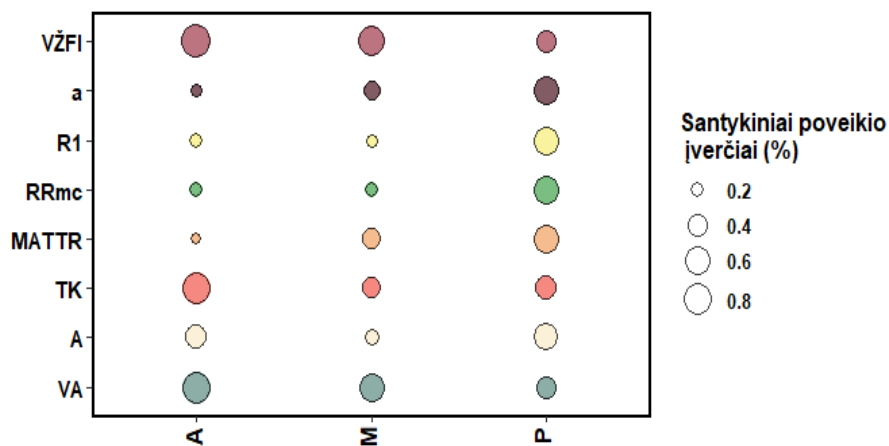
Didesnė aktyvumo raiška yra būdingesnė publicistiniams tekstams (santykinio poveikio įverčio reikšmė – 0,54875), nors administraciniai tekstai šio požymio raiška ne itin daug jiems nusileidžia (santykinio poveikio įverčio reikšmė – 0,41830). Kita vertus, didesnė aktyvumo raiška publicistiniuose tekstuose nėra itin ryški: viso labo tik ~55 % tikimybė, kad didesnė aktyvumo indikatoriaus reikšmė priklausys šio FS tekstui. Aiškiai mažesnė aktyvumo raiška būdingesnė moksliniams (santykinio poveikio įverčio reikšmė – 0,22908). Tad tai rodo, jog publicistiniams tekstams būdingesnis didesnis (nors ši tikimybė ir nėra itin didelė), lyginant su būdvardžiais, veiksmažodžių kiekis, o moksliniams – mažesnis. Administraciniai tekstai (santykinio poveikio įverčio reikšmė – 0,41830) pagal šį požymį yra artimesni publicistiniams negu moksliniams.

Galiausiai sintaksiškai sudėtingesni tekstai arba didesnės indikatoriaus VA vertės, rodančios vidutinį atstumą tarp dviejų iš eilės einančių veiksmažodžių tekste, yra būdingesni AST (santykinio poveikio įverčio reikšmė – 0,77809). Sintaksiškai paprastesni tekstai (mažesnės VA vertės) tipiškesni PST (santykinio poveikio įverčio reikšmė – 0,39670). Moksliniai tekstai pagal šį požymį

yra artimesni administraciniams tekstams nei publicistiniams (santykinio poveikio įverčio reikšmė – 0,60010).

5.2 Funkcinių stilių tarpusavio panašumų ir skirtumų apibendrinimas

Labiausiai tiriamieji FS apskritai tarpusavyje skiriasi pagal informacijos srautą, kitaip – pagal vidutinį skirtingų žodžių ir žodžių formų kiekį tekste. Teksto suprantamumo sudėtingumas (arba vidutinis žodžių ir žodžių formų ilgis tekste), teminių / turinio žodžių ir žodžių formų įvairovė taip priklauso prie tiriamuosius FS geriausiai skiriančių požymių, nors jų gebėjimas diferencijuoti FS yra mažesnis nei informacijos srauto. Mažiausiai – pagal aktyvumo ir deskriptyvumo raišką. Kitaip sakant, veiksmažodžių ir būdvardžių santykis nėra geriausias skiriamasis AST, MST ir PST požymis. Toliau detalizuojami šie skirtumai. Vizualiai šie tiriamųjų FS požymiai, atsiremiant į santykinio poveikio įverčius (detaliau žr. 34 lentelę), pateikiami 17 pav.



17 pav. FS požymius apibendrinanti diagrama (A – administracinis FS, M – mokslinis FS, P – publicistinis FS)

Administraciniai ir publicistiniai tekstai tarpusavyje labiausiai skiriasi pagal teksto suprantamumo sudėtingumą (arba vidutinis žodžių ir žodžių formų ilgį tekste), labai dažnų žodžių ir žodžių formų (paprastai – funkcinių) kiekį, informacijos srautą (arba vidutinį skirtingų žodžių ir žodžių formų kiekį tekste) ir sintaksinės struktūros sudėtingumą (arba vidutinį atstumą žodžiais tarp 2 iš eilės einančių veiksmažodžių). Moksliniai ir publicistiniai tekstai

labiausiai skiriasi pagal teminių / turinio žodžių ir žodžių formų įvairovę, žodžių ir žodžių formų koncentraciją (arba žodžių ir žodžių formų, turinčių didelį dažnį tekste, kiekį) bei dinamiškumą (veiksmožodžių ir būdvardžių santykį). Administraciniai ir moksliniai tekstai labiausiai skiriasi pagal teminės koncentracijos raišką (turinio / teminių žodžių, dažnumų bemaž prilygstančių funkciniams, kiekį), o panašūs – pagal žodžių ir žodžių formų koncentraciją (arba žodžių ir žodžių formų, turinčių didelį dažnį tekste, kiekį). Toliau tekste detalizuojamos funkcinių stilių panašumų-skirtumų pagal tyrime naudotus kiekybinius indikatorius kryptys, t. y. indikatorių raiškos būdingumo tendencijas.

Administraciniams tekstams būdingesni vidutiniškai ilgesnis žodžių ir žodžių formų ilgis (kitaip – suprantamumo prasme sudėtingesni tekstai), didesnis kiekis labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų. Be to, šiems tekstams tipiškesnė mažesnė turinio / teminių žodžių ir žodžių formų įvairovė, nors pagal šią tendenciją skirtumas tarp administracinių ir mokslinių tekstų yra nedidelis, bet vis tik statistiškai reikšmingas. Administraciniams tekstams taip pat būdingesnė mažesnė žodžių ir žodžių formų koncentracija. Pagal šią tendenciją administraciniai ir moksliniai tekstai yra panašūs. Administraciniams tekstams būdingesnis ir mažesnis informacijos srautas (arba vidutiniškai mažesnis kiekis skirtingų žodžių ir žodžių formų). Kita vertus, šiems tekstams tipiškesnė didesnė teminė koncentracija (didesnis kiekis teminių / turinio žodžių ir žodžių formų, dažnumu nedaug nusileidžiančių funkciniams žodžiams ir žodžių formoms). Pagal dinamiškumo raišką administraciniai tekstai yra panašesni į publicistinius tekstus, t. y. šioms tekstų grupėms būdingesnis didesnis veiksmožodžių kiekis lyginant su būdvardžiais negu moksliniuose tekstuose. Galiausiai administraciniams tekstams būdingesnė kompleksiškesnė sintaksinė struktūra (kitaip – didesnis atstumas žodžiais tarp 2 iš eilės einančių veiksmožodžių).

Moksliniai tekstai suprantamumo sudėtingumo atžvilgiu (kitaip – pagal vidutinį žodžių ir žodžių formų ilgį) yra artimesni administraciniams negu publicistiniams, t. y. tikėtiniesni suprantamumo prasme sudėtingesni tekstai. Pagal labai dažnų žodžių ir žodžių formų kiekį moksliniai tekstai taip pat yra artimesni administraciniams, kuriems būdingesnis didesnis kiekis labai dažnų žodžių ir žodžių formų kiekis. MST tekstams tipiškesnė mažesnė turinio / teminių žodžių įvairovė. O pagal žodžių ir žodžių formų koncentraciją tarp šių ir administracinių tekstų statistiškai reikšmingo skirtumo nerasta. Informacijos srauto požiūriu moksliniai tekstai krypta ir vėl administracinių tekstų, kuriems būdingesnis mažesnis informacijos srautas, link. Kita vertus, MST tekstams tipiškesnė mažesnė teminė koncentracija bei mažesnė

dinamiškumo raiška. Sintaksinės struktūros kompleksškumo atžvilgiu moksliniai tekstai artimesni administraciniams, kuriems tipiškesni šiuo požiūriu sudėtingesni tekstai.

Publicistiniams tekstams būdingesni vidutiniškai trumpesni žodžiai ir žodžių formos arba, kitaip, suprantamumo prasme paprastesni tekstai. Šiems tekstams taip pat tipiškesnis mažesnis labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų kiekis bei didesnė turinio / teminių žodžių ir žodžių formų įvairovė. Tuo pat metu publicistiniams tekstams būdingesnė ir didesnė žodžių ir žodžių formų koncentracija, t. y. kai tekste esama keleto žodžių ir žodžių formų, turinčių itin didelius dažnius, o likusių žodžių ir žodžių formų dažnis tekste yra žymiai mažesnis. Be to, šiems tekstams tipiškesnis didesnis informacijos srautas (arba vidutiniškai didesnis kiekis skirtingų žodžių ir žodžių formų). Teminės koncentracijos požiūriu PST tekstai artimesni moksliniams (būdingesnė mažesnė teminė koncentracija), o dinamiškumo atžvilgiu – administraciniams (tipiškesnis didesnis dinamiškumas arba didesnis kiekis veiksmožodžių lyginant su būdvardžiais). Galiausiai publicistiniams tekstams būdingesnė paprastesnė sintaksinė struktūra – mažesnis atstumas (žodžių formomis) tarp 2 iš eilės einančių veiksmožodžių.

Neparametriniai išvadų statistikos testai leido įvertinti FS kalbinės raiškos požymių, kuriuos žymi statistiniai indikatoriai, mastą ir palyginti, kaip ir kuo tiksliniai FS tarpusavyje yra panašūs bei kaip ir kuo – skiriasi. Vizualiai šie tiriamųjų FS panašumai ir skirtumai pateikti 17 pav. Šie apibendinti FS kalbinės raiškos panašumai ir skirtumai leido išskirti tiriamųjų FS modelius pagal indikatorių raiškos mastą – tai ir buvo pristatyta šiame skyriuje. Šie modeliai leidžia geriau suprasti FS charakteristikas, tad gauti rezultatai gali prisidėti kuriant automatines tekstų paieškos ar klasifikavimo sistemas. Indikatorių kaip požymių naudojimas kartu su mašininio mokymo metodais buvo tinkamas pasirinkimas tekstų klasifikavimui į FS, kaip parodė disertacijos autorės kartu su bendraautoriais publikuotas tyrimas¹¹⁸, kuris dėl disertacijos apimties apribojimų į ją nebuvo įtrauktas.

¹¹⁸ Mandravickaitė, J., Krilavičius, T., & Man, K. L. (2019). Document Classification to Functional Styles (Domains of Use): Lithuanian Case. *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems*, 8(1), 38-41.

IŠVADOS

Kaip jau buvo minėta, šio darbo tikslas – kiekybinės lingvistikos metodais ištirti 3 lietuvių kalbos funkcinių stilių – administracinio, mokslinio ir publicistinio – kalbinę raišką ir įvertinti šių funkcinių stilių panašumus ir skirtumus pagal kalbinės raiškos požymių (teksto sudėtingumo (žymi indikatorius VŽFI), labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (žymi indikatorius a), žodžių ir žodžių formų (retesnių) įvairovės (žymi indikatorius R_1), žodžių ir žodžių formų koncentracijos (žymi indikatorius RR_{mc}), informacijos srauto (žymi indikatorius MATTR), teminės koncentracijos (žymi indikatorius TK), dinamiškumo (žymi indikatorius A), deskriptyvumo (žymi indikatorius D) ir teksto sintaksinio kompleksiskumo (žymi indikatorius VA)) mastą, taip išskiriant apibendrintus tikslinių funkcinių stilių kalbinės raiškos modelius. Juos atskleidžia tyrimo išvadų visuma. Atliktas rašytinių tekstų funkcinių stilių – administracinio, mokslinio ir publicistinio – kalbinės raiškos tyrimas taikant kiekybinės lingvistikos metodus leidžia daryti tokias pagrindines išvadas:

1. Pasirinkti indikatoriai VŽFI, a , R_1 , RR_{mc} , MATTR, TK, A, D ir VA leido įvertinti lietuvių kalbos administracinio, mokslinio ir publicistinio funkcinių stilių kalbinę raišką, tad prielaida apie jų tinkamumą kalbinės raiškos tyrimui iš kiekybinės perspektyvos pasitvirtino.
2. Nustatyti atskirų tiriamųjų funkcinių stilių kalbinę raišką geriausiai reprezentuojančių požymių kombinacijos:
 - 2.1 administracinio funkcinio stiliaus tekstus analizuojant atskirai nuo mokslinio ir publicistinio funkcinių stilių, šio stiliaus kalbinę raišką geriausiai reprezentuojančių požymių rinkinį (požymiai išdėstyti pagal bendrą procentinį indėlį, pradedant nuo didžiausio) sudaro:
 - žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovė (žymi indikatorius R_1);
 - aktyvumas / dinamiškumas, kitaip – veiksmažodžių ir būdvardžių santykis (žymi indikatorius A);
 - deskriptyvumas / aprašomumas, kitaip – būdvardžių ir veiksmažodžių santykis (žymi indikatorius D);
 - žodžių ir žodžių formų koncentruotumas, kitaip – itin didelio dažnio žodžių ir žodžių formų kiekis tekste (žymi indikatorius RR_{mc});
 - informacijos srautas, kitaip – skirtingų žodžių ir žodžių formų kiekis tekste (žymi indikatorius MATTR);

- teminė koncentracija, kitaip – iš esmės retesnių, pagrindinėms kalbos dalims priklausančių, tačiau tekste labai dažniems (funkciniams) dažnumu bemaž prilygstančių žodžių ir žodžių formų kiekis (žymi indikatorius TK).

2.2 Mokslinio funkcinio stiliaus tekstus analizuojant atskirai nuo administracinio ir publicistinio funkcinio stilių, šio stiliaus kalbinę raišką geriausiai reprezentuojančių požymių rinkinį (požymiai išdėstyti pagal bendrą procentinį indėlį, pradedant nuo didžiausio) sudaro:

- aktyvumas / dinamiškumas, kitaip – veiksmažodžių ir būdvardžių santykis (žymi indikatorius A);
- deskriptyvumas / aprašomumas, kitaip – būdvardžių ir veiksmažodžių santykis (žymi indikatorius D);
- žodžių ir žodžių formų koncentruotumas, kitaip – itin didelio dažnio žodžių ir žodžių formų kiekis tekste (žymi indikatorius RR_{mc});
- žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovė (žymi indikatorius R_1);
- teksto sintaksinės struktūros kompleksiskumas – atstumas žodžiais tarp dviejų iš eilės einančių veiksmažodžio formų (žymi indikatorius VA).

2.3 Publicistinio funkcinio stiliaus tekstus analizuojant atskirai nuo administracinio ir mokslinio funkcinio stilių, šio stiliaus kalbinę raišką geriausiai reprezentuojančių požymių rinkinį (požymiai išdėstyti pagal bendrą procentinį indėlį, pradedant nuo didžiausio) sudaro:

- žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovė (žymi indikatorius R_1);
- teksto sintaksinės struktūros kompleksiskumas – atstumas žodžiais tarp dviejų iš eilės einančių veiksmažodžio formų (žymi indikatorius VA);
- aktyvumas / dinamiškumas, kitaip – veiksmažodžių ir būdvardžių santykis (žymi indikatorius A);
- deskriptyvumas / aprašomumas, kitaip – būdvardžių ir veiksmažodžių santykis (žymi indikatorius D);
- teminė koncentracija, kitaip – iš esmės retesnių, pagrindinėms kalbos dalims priklausančių, tačiau tekste labai dažniems (funkciniams) dažnumu bemaž prilygstančių žodžių ir žodžių formų kiekis (žymi indikatorius TK);

- žodžių ir žodžių formų koncentruotumas, kitaip – itin didelio dažnio žodžių ir žodžių formų kiekis tekste (žymi indikatorius RR_{mc}).
3. Nustatyti panašumai ir skirtumai tarp tiriamųjų funkcinių stilių:

3.1. Administraciniam stiliui, lyginant visus tiriamuosius funkcinius stilius, būdingiau:

- ilgesnis vidutinis žodžių ir žodžių formų ilgis (kitaip – suprantamumo prasme sudėtingesni tekstai; žymi indikatorius $V\check{Z}FI$);
- didesnė gausa labai dažnų (paprastai – funkcinių, bet nebūtinai) žodžių ir žodžių formų (žymi indikatorius a);
- mažesnė žodžių ir žodžių formų koncentracija – pagal šį rodiklį administraciniai ir moksliniai tekstai yra panašūs (indikatorius RR_{mc});
- mažesnis informacijos srautas (arba vidutiniškai mažesnis kiekis skirtingų žodžių ir žodžių formų; žymi indikatorius $MATTR$);
- didesnė teminė koncentracija (didesnis kiekis iš esmės retesnių, pagrindinėms kalbos dalims priklausančių, tačiau administraciniuose tekstuose labai dažniems (funkciniams) dažnumu bemaž prilygstančių žodžių ir žodžių formų);
- kompleksiškesnė sintaksinė struktūra (kitaip – didesnis atstumas žodžiais tarp 2 iš eilės einančių veiksmažodžių; žymi indikatorius VA);
- pagal žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovės raišką administraciniai tekstai yra panašesni į mokslinius – būdingesnė mažesnė retesnių žodžių ir žodžių formų įvairovė (žymi indikatorius R_1);
- pagal aktyvumo (veiksmažodžių ir būdvardžių santykį tekste; žymi indikatorius A) raišką administraciniai tekstai yra panašesni į publicistinius (būdingesnė didesnė aktyvumo raiška) tekstus nei į mokslinius.

3.2 Mokslinis stilius pagal dalį kalbinės raiškos požymių yra tarp administracinio ir publicistinio. Tad šiais atvejais mokslinių tekstų kalbinė raiška įvertinama pagal jos artimumą priešingų polių – administracinio ir publicistinio – stilių tekstams:

- moksliniams tekstams tipiškesnė mažesnė žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovė (žymi indikatorius R_1);
- moksliniams tekstams tipiškesnė mažesnė teminė koncentracija, t. y. mažesnis kiekis iš esmės retesnių, pagrindinėms kalbos dalims priklausančių, tačiau moksliniuose tekstuose labai dažniems (funkciniams) dažnumu bemaž prilygstančių žodžių ir žodžių formų;
- moksliniams tekstams būdingesnė mažesnė aktyvumo raiška (žymi indikatorius A), kitaip – lyginant su būdvardžiais, mažesnis kiekis veiksmazodžių;
- suprantamumo sudėtingumo atžvilgiu (kitaip – pagal vidutinį žodžių ir žodžių formų ilgį; žymi indikatorius $V\check{Z}FI$) moksliniai tekstai yra artimesni administraciniams (būdingesni suprantamumo atžvilgiu sudėtingesni tekstai – vidutiniškai ilgesni žodžiai ir žodžių formos) negu publicistiniams (būdingesni suprantamumo atžvilgiu paprastesni tekstai – vidutiniškai trumpesni žodžiai ir žodžių formos);
- mažesnė žodžių ir žodžių formų koncentracija (žymi indikatorius RR_{mc}) – pagal šį rodiklį administraciniai ir moksliniai tekstai yra panašūs;
- pagal labai dažnų žodžių ir žodžių formų kiekį (indikatorius a) moksliniai tekstai taip pat yra artimesni administraciniams (būdingesnė didesnė labai dažnų žodžių ir žodžių formų gausa);
- pagal informacijos srautą (vidutinį skirtingų žodžių ir žodžių formų kiekį tekste; žymi indikatorius $MATTR$) moksliniai tekstai artimesni administraciniams (būdingesnis didesnis informacijos srautas – didesnis vidutinis skirtingų žodžių ir žodžių formų kiekis);
- pagal sintaksinės struktūros kompleksiško rodiklį (indikatorius VA) moksliniai tekstai artimesni administraciniams (tekstams būdingesnė kompleksiškesnė sintaksė).

3.3. Publicistiniam stiliui, lyginant visus tiriamuosius funkcinis stilius, būdingiau:

- vidutiniškai trumpesni žodžiai ir žodžių formos arba, kitaip, suprantamumo prasme paprastesni tekstai;

- mažesnė labai dažnų (paprastai – funkcinių) žodžių ir žodžių formų gausa (žymi indikatorius a);
 - didesnė žodžių ir žodžių formų (retesnių – priklausančių pagrindinėms kalbos dalims) įvairovė;
 - didesnis žodžių ir žodžių formų koncentruotumas, t. y. mažesnis kiekis žodžių ir žodžių formų, turinčių itin didelius dažnius;
 - didesnis informacijos srautas (arba vidutiniškai didesnis kiekis skirtingų žodžių ir žodžių formų);
 - paprastesnė sintaksinė struktūra – mažesnis atstumas tarp 2 iš eilės einančių veiksmažodžių;
 - teminės koncentracijos požiūriu publicistiniai tekstai artimesni moksliniams (būdingesnė mažesnė teminė koncentracija, kitaip – mažesnis kiekis iš esmės retesnių, pagrindinėms kalbos dalims priklausančių, tačiau moksliniuose tekstuose labai dažniems (funkciniams) dažnumu bemaž prilygstančių žodžių ir žodžių formų);
 - aktyvumo (veiksmažodžių ir būdvardžių santykis tekste; žymi indikatorius A) atžvilgiu publicistiniams tekstams būdingesni dinamiškesni tekstai, nors ši tendencija nėra labai stipri, nes administracinių tekstų aktyvumo raiška mastu nedaug nusileidžia publicistiniams.
4. Nustatyti ryšiai tarp tiriamųjų požymių:
- 4.2 Labai stiprus ryšys visiems trimis tiriamiesiems funkciniams stiliams nustatytas tarp:
- žodžių ir žodžių formų (retesnių, priklausančių pagrindinėms kalbos dalims) įvairovės ir žodžių ir žodžių formų koncentruotumo (indikatoriai R_1 ir RR_{mc});
 - dinamiškumo ir deskriptyvumo (indikatoriai A ir D).
- 4.3 Vidutinio stiprumo ryšiai nustatyti:
- 4.3.1 Administraciniam stiliui – 8 požymių poroms:
- žodžių ir žodžių formų (retesnių) įvairovės ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai R_1 ir a); žodžių ir žodžių formų koncentracijos ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai RR_{mc} ir a); informacijos srauto ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai $MATTR$ ir a); informacijos srauto ir žodžių ir žodžių formų (retesnių) įvairovės (indikatoriai $MATTR$ ir R_1); informacijos srauto ir žodžių ir žodžių formų koncentracijos

(indikatoriai MATTR ir RR_{mc}); informacijos srauto ir sintaksinio kompleksiško (indikatoriai MATTR ir VA);

- dinamiškumo ir sintaksinio kompleksiško (indikatoriai A ir VA); deskriptyvumo ir sintaksinio kompleksiško (indikatoriai D ir VA).

4.3.2 Moksliniam stiliui – 10 požymių porų:

- teksto sudėtingumo ir žodžių ir žodžių formų (retesnių) įvairovės (indikatoriai VŽFI ir R_1); teksto sudėtingumo ir dinamiškumo (indikatoriai VŽFI ir A); teksto sudėtingumo ir deskriptyvumo (indikatoriai VŽFI ir D); žodžių formų (retesnių) įvairovės ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai R_1 ir a); dinamiškumo ir sintaksinio kompleksiško (indikatoriai A ir VA); dinamiškumo ir teminės koncentracijos (indikatoriai A ir TK); deskriptyvumo ir sintaksinio kompleksiško (indikatoriai D ir VA); deskriptyvumo ir teminės koncentracijos (indikatoriai D ir TK); informacijos srauto ir teminės koncentracijos (indikatoriai MATTR ir TK); teksto sudėtingumo ir teminės koncentracijos (indikatoriai VŽFI ir TK).

4.3.3 Publicistiniam stiliui – 11 požymių porų:

- teksto sudėtingumo ir teminės koncentracijos (indikatoriai VŽFI ir TK); teksto sudėtingumo ir dinamiškumo (indikatoriai VŽFI ir A); teksto sudėtingumo ir deskriptyvumo (indikatoriai VŽFI ir D); teksto sintaksinio kompleksiško ir teksto sudėtingumo (indikatoriai VA ir VŽFI); žodžių formų (retesnių) įvairovės ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai R_1 ir a); žodžių ir žodžių formų koncentracijos ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai RR_{mc} ir a); informacijos srauto ir labai dažnų ir retesnių žodžių ir žodžių formų santykio tekste (indikatoriai MATTR ir a); informacijos srauto ir žodžių ir žodžių formų (retesnių) įvairovės (indikatoriai MATTR ir R_1); informacijos srauto ir žodžių ir žodžių formų koncentracijos (indikatoriai MATTR ir RR_{mc}); dinamiškumo ir sintaksinio kompleksiško (indikatoriai A ir VA); deskriptyvumo ir sintaksinio kompleksiško (indikatoriai D ir VA).

Šiame darbe atliktas administracinio, mokslinio ir publicistinio funkcinių stilių kalbinės raiškos tyrimas leido ištirti šių funkcinių stilių raišką tiek kiekvieno atskirai, tiek lyginant ir priešpriešinant vieną su kitu, taip išskiriant apibendrintus tikslinių funkcinių stilių kalbinės raiškos modelius. Pasiūlyta metodika kalbinei raiškai tirti didelės apimties tekstynuose kiekybiniu aspektu tiko tiek atskirų tekstų grupių (atskirų funkcinių stilių) būdingų požymių, tiek tekstų grupių (funkcinių stilių) tarpusavio panašumų ir skirtumų pagal kiekybinių požymių raiškos mastą įvertinimui.

Kita vertus, tyrime esama ir **ribotumų**. Lietuvių kalboje tradiciškai išskiriami 5 funkciniai stiliai, o šiame tirti tik 3. Analizuojant visą funkcinių stilių sistemą sudarančių stilių kalbinės raiškos požymių dinamiką (stilių tarpusavio panašumai ir skirtumai, jų mastas) rezultatai gali šiek tiek pasikeisti. Taip pat paminėtina, kad tyrime publicistikos stilių reprezentavo naujienų portalo Delfi tekstai, kurie, nors ir įvairialypiai, nebūtinai aprėpia specializuotą, laikraščių ar regioninę publicistiką. Be to, kalbinę raišką sudaro įvairių kalbos lygmenų požymių įvairovė, o šiame tyrime naudoti tik 9, kurių nors ir pakako vieno funkcinio stiliaus atskyrimui nuo kito, tačiau neapima visos kalbinės raiškos.

Atlikus šioje disertacijoje aprašomą tyrimą manoma, kad **tekstų tyrėjams gali būti naudingi šie pastebėjimai:**

- Funkciniai stiliai yra bendros funkcinių stilių sistemos dalis, tad juos ir tirti būtų naudingiau ne po vieną, o lyginant juos tarpusavyje. Tai leistų tirti požymių raiškos mastą ir dinamiką, užfiksuoti funkcinių stilių kitimą, ryšius, jų vidinę diferenciaciją.
- Tiriant funkcinių stilių (ar kita logika grupuojamų tekstų) kalbinę raišką naudinga remtis kombinacija, o ne pavieniu požymiu. Tai leistų tiksliau įvertinti kalbinės raiškos mastą, kadangi požymiai dažnai yra susiję tarpusavyje.

Šio tyrimo rezultatai galėtų būti panaudojami automatiniam tekstų klasifikavimui¹¹⁹ ir kategorizavimui – tekstų automatinis suskirstymas pagal funkcinis stilius gali būti aktualus bibliotekoms, knygynams, kalbinių

¹¹⁹ Kad statistinius indikatorius galima sėkmingai panaudoti automatiniam tekstų klasifikavimui, parodė į disertaciją neįtrauktas tyrimas, disertacijos autorės publikuotas kartu su bendraautorais: Mandravickaitė, J., Krilavičius, T., & Man, K. L. (2019). Document Classification to Functional Styles (Domains of Use): Lithuanian Case. *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems*, 8(1), 38-41.

resursų (tekstynų, duomenų bazių ir pan.) planavimui ir kūrimui, internete egzistuojančių tekstų gausos įvertinimui, informacijos paieškai pagal vartotojo pateiktus kriterijus, dokumentų valdymo sistemoms (dokumentų/tekstų rekomendacijoms, dokumentų atitikimas saugumo reikalavimams ir pan.) bei panašioms tikslams. Be to, funkcinį stilių kalbinės raiškos tyrimo rezultatai gali būti panaudojami automatinio tekstų generavimo srityje, pavyzdžiui, automatiniam prekių/produktų aprašymų generavimui iš duomenų bazės įrašo ar paveikslėlio. Šio tyrimo rezultatai taip pat sudarytų prielaidas tekstų stiliaus automatiniam keitimui iš vieno į kitą – šis taikymas galėtų būtų naudingas verslui, kai vieno stiliaus tekstas, pavyzdžiui, administracinio, gali būti automatiškai konvertuotas į kito stiliaus, pavyzdžiui, publicistinio, taip taupant laiką, finansinius ir žmogiškuosius resursus.

LITERATŪROS SĄRAŠAS

- Altmann, G. (1980). Prolegomena to Menzerath's law. *Glottometrika*, 2(2), 1—10.
- Altmann, G. (1983). A Law of Change in Language. *Quantitative Linguistics*, 18, 104—115.
- Altmann, G. (1985). On the dynamic approach to language. Linguistic dynamics. *Discourses, procedures and evolution*, 181—189.
- Altmann, G. (1993). Science and linguistics. *Contributions to quantitative linguistics*, 3—10. Springer.
- Altmann, G. (2002a). *Einführung in die quantitative Lexikologie* (Bd. 5). Göttingen: Peust und Gutschmidt.
- Altmann, G. (2002b). Zipfian linguistics. *Glottometrics*, 3(2002), 19—26.
- Altmann, G. (2016). Types of Hierarchies in Language. *Glottometrics*, 34, 44—55.
- Altmann, G., Köhler, R. (1996). Language Forces and synergetic modelling of language phenomena. *Glottometrika* 15, 62—76.
- Altmann, G., Köhler, R. (2015). *Forms and degrees of repetition in texts: detection and analysis* (Vol. 68). Walter de Gruyter GmbH & Co KG.
- Andreev, S., Popescu, I.-I., Altmann, G. (2017). Motifs of adnominals in Russian. *Glottometrics*, 38, 77—106
- Argamon, S., Koppel, M., Fine, J., Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text-the Hague then Amsterdam then Berlin-*, 23, 321—346.
- Basharin, G. P., Langville, A. N., ir Naumov, V. A. (2004). The life and work of AA Markov. *Linear algebra and its applications*, 386, 3—26.
- Beliankou, A., Köhler, R. (2010). The distribution of parts-of-speech in Russian texts. *Glottometrics*, 20, 59—69.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., Matsuo, A. (2018). An R package for the quantitative analysis of textual data. *J. Open Source Software*, 3(30), 774.
- Bernhard, G., Altmann, G. (2008). The phoneme-grapheme relationship in Italian. *Analyses of Script. Properties of Characters and Writing Systems*, 13—23.
- Best, K.-H. (2006). Wilhelm Förstemann (1822–1906). *Glottometrics* 12, 77—86.
- Best, K.-H. (2007). Otto Behaghel (1854-1936). *Glottometrics* 14, 80—86.
- Best, K.-H. (2006). XVI. Ernst Wilhelm Förstemann (1822-1906). *Glottometrics*, 12, 77—86.

- Best, K.-H. (2007). XXX. Gustav Herdan (1897-1968). *Glottometrics*, 15, 92—96.
- Best, K.-H., Altmann, G. (1986). Untersuchungen zur Gesetzmäßigkeit von Entlehnungsprozessen im Deutschen. *Folia Linguistica Historica*, 20 (Historica vol. 7, 1), 31—42.
- Best, K.-H., Altmann, G. (2018). Word Length with G. Herdan. *Glottometrics*, 42, 86—90.
- Byrne, B. M. (2010). *Multivariate applications series. Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York, NY, US: Routledge/Taylor ir Francis Group.
- Bitinienė, A. (1983). *Mokslinis stilius*. Vilnius.
- Bitinienė, A. (1997). *Funkciniai stiliai: sakinio ilgis ir struktūra*. Vilnius: VPU 1-kl.
- Bitinienė, A. (1998). Administracinis stilius ir jo sakinio ilgis. *Kalbotyra*, 47(1), 17—27.
- Bitinienė, A. (2000). Vientisiniai mokslinio stiliaus sakiniai. *Kalbotyra*, 48, 19—27.
- Bitinienė, A. (2001). Grožinio stiliaus prozos tekstų tiesioginės kalbos sakinio ilgis ir struktūra. *Kalbotyra*, 50, 17—28.
- Bitinienė, A. (2002). Administracinio stiliaus tekstų sakinio ilgio ir struktūros pokyčiai. *Filologija*, 7, 4—8.
- Bitinienė, A. (2003). Lietuvių ir latvių kalbų publicistinio stiliaus tekstų sakinio ilgis. *Lituanistica*, 3, 85—95.
- Bitinienė, A. (2004). Klausiamieji sakiniai publicistinio stiliaus tekstuose. *Kalbotyra*, 1(53), 5—12.
- Bitinienė, A. (2006). Juozo Pikčilingio funkcinė stilistika. *Kalbos kultūra*, 299—306.
- Bitinienė, A. (2007a). Mokomųjų tekstų sakinio ilgis ir struktūra. *Žmogus ir žodis*, 9(1), 12—18.
- Bitinienė, A. (2007b). Mokslinio stiliaus bendrojo ir individualiojo komponento santykis. *Acta humanitarica universitatis Saulensis*, 3, 32—39.
- Bitinienė, A. (2007c). *Publicistinis stilius*. Vilnius: Vilniaus pedagoginio universiteto leidykla.
- Bitinienė, A. (2009). Sakinių siejimo priemonių vartojimas mokslinio stiliaus tekstuose. *Žmogus ir žodis*, 11(1), 21—26.
- Bružaitė, N., Rekašius, T. (2016). Žodžių dažnių pasiskirstymo analizė skirtingų žanrų lietuvių kalbos tekstuose. *Lithuanian Journal of Statistics*, 55(1), 61—69.

- Buk, S. N., Rovenchak, A. A. (2004). Rank-Frequency analysis for functional style corpora of Ukrainian. *Journal of Quantitative Linguistics*, 11(3), 161—171.
- Bulmer, M. G. (1979). *Principles of statistics*. Courier Corporation.
- Bumbulienė, I., Mandravickaitė, J., Boizou, L., ir Krilavičius, T. (2017). An overview of Lithuanian Internet media n-gram corpus. *CEUR Workshop proceedings [electronic resource]: SYSTEM 2017: proceedings of the symposium for Young Scientists in Technology, Engineering and Mathematics, Kaunas, Lithuania, April 28, 2017*. Aachen: CEUR-WS, 2017, Vol. 1853, 24—28.
- Butkutė, L. (2008). Okazionaliųjų frazeologizmų stilistinės išgalės publicistiniame stiliuje. *Respectus Philologicus*, 13(18), 199—209.
- Butkutė, L. (2010). *Okazionaliųjų frazeologizmų stilistinės išgalės publicistiniame stiliuje*. Daktaro disertacija, Vilniaus universitetas.
- Canning, P. (2017). Functionalist stylistics. *The Routledge Handbook of Stylistics*, 63—85, Routledge.
- Casas, P. (2016). funModeling: Learning Data Cleaning, Visual Analysis and Model Performance (version 1.0). Nuskaityta iš <https://CRAN.R-project.org/package=funModeling>
- Chen, R., Liu, H. (2014). Quantitative aspects of journal of quantitative linguistics. *Journal of Quantitative Linguistics*, 21(4), 299—340.
- Chen, R., Liu, H. (2015). Ideologies of supreme court justices: Quantitative thematic analysis of multiple opinions of “Bush V. Gore 2000”. *Glottology*, 6, 299—322.
- Chen, R., Liu, H. (2018). Thematic Concentration as a Discriminating Feature of Text Types. *Journal of Quantitative Linguistics*, 25, 53—76.
- Chen, R., Liu, H., ir Altmann, G. (2016). Entropy in different text types. *Digital Scholarship in the Humanities*, 32, 528—542.
- Chloupek, J. (1993a). Language Varieties and Styles in Communication. *Studies in Functional Stylistics*, 36, 68—91.
- Chloupek, J. (1993b). Publicist Style. *Studies in Functional Stylistics*, 36, 112—126.
- Chovanec, J. (2014). Approaching Czech linguistic functionalism. *Chapters from the history of Czech functional linguistics*, 6—26.
- Collinge, N. E. (1985). *The laws of Indo-European* (T. Vol. 35). John Benjamins Publishing.
- Condon, E. U. (1928). Statistics of vocabulary. *Science*, 67, 300.
- Coupland, N. (2007). *Style: Language variation and identity*. Cambridge University Press.

- Covington, M. A., McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of quantitative linguistics*, 17, 94—100.
- Čech, R. (2011). Frequency structure of New Year's presidential speeches in Czech. The authorship analysis. *Issues in Quantitative Linguistics*, 2, 82—94.
- Čech, R. (2015). Text length and the lambda frequency structure of a text. *Sequences in Language and Text*, 69, 71—88.
- Čech, R. (2016). *Tematická koncentrace textu v češtině*. ÚFAL, Ústav formální a aplikované lingvistiky.
- Čech, R., Garabík, R., Altmann, G. (2015). Testing the thematic concentration of text. *Journal of Quantitative Linguistics*, 22, 215—232.
- Čech, R., Popescu, I.-I., ir Altmann, G. (2014). *Metody kvantitativní analýzy (nejen) básnických text*. Univerzita Palackého v Olomouci.
- Čekanavičius, V. ir Murauskas, G. (2001). *Statistika ir jos taikymai (I knyga)*: [vadovėlis aukštųjų mokyklų studentams]. Vilnius: TEV.
- Čekanavičius, V., ir Murauskas, G. (2002). *Statistika ir jos taikymai (II knyga)*: [vadovėlia aukštųjų mokyklų studentams]. Vilnius: TEV.
- Česnulienė, V. (2012). Gramatinė kohezija mokslinio ir publicistinio stiliaus tekstuose. *Filologija*, 17, 45—55.
- Čičirkaitė, R. (2005). Lingvistiniai reklamos tekstų požymiai. *Žmogus ir žodis*, 77(1), 73—77.
- Daneš, F. (1987). On Prague School Functionalism in Linguistics. R. Dirven, ir V. Fried (Eds.), *Functionalism in Linguistics*, 3—38. John Benjamins Publishing.
- Daneš, F. (1994). Prague school functionalism as a precursor of text linguistics. *Cahiers de l'ILSL*, 5, 117—126.
- Dinno, A. (2015). Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. *The Stata Journal*, 15(1), 292—300.
- Dinu, A., Dinu, L. (2009). On the behavior of Romanian syllables related to minimum effort laws. *Proceedings of the Workshop Multilingual resources, technologies and evaluation for central and Eastern European languages*, 9—13.
- Doležel, L. (1968). Russian and Prague School functional stylistics. *Style*, 2(2), 143—158.
- Dormann, C. F. (2017). *Parametrische Statistik*. Springer.
- Elderton, W. (1949). A few statistics on the length of English words. *Journal of the Royal Statistical Society. Series A (General)*, 112(4), 436—445.
- Ellis, A. R., Burchett, W. W., Harrar, S. W., Bathke, A. C. (2017). Nonparametric inference for multivariate data: the R package nrmv. *Journal of Statistical Software*, 76(4), 1—18.

- Emmott, C., Alexander, M. (2015). Defamiliarization and foregrounding. *The Bloomsbury companion to stylistics*, 289—307.
- Estoup, J.-B. (1912). *Gammes sténographiques*. Recueil de textes choisis pour l'acquisition méthodique de la vitesse, précédé d'une introduction par J.-B. EstouParis: Institut sténographique.
- Everitt, B., Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer Science & Business Media.
- Fan, F., Altmann, G. (2008). On meaning diversification in English. *Glottometrics*, 17, 66—78.
- Fan, F., Grzybek, P., Altmann, G. (2010). Dynamics of word length in sentence. *Glottometrics*, 20, 70—109.
- Fang, Y., Liu, H. (2015). Comparison of vocabulary richness in two translated Honglouneng. *Glottometrics*, 31, 54—75.
- Field, A., Miles, J., Field, Z. (2012). *Discovering statistics using R*. Sage publications.
- Förstemann, E. (1852). Numerische lautverhältnisse im griechischen, lateinischen und deutschen. *Zeitschrift für vergleichende Sprachforschung auf dem Gebiete des Deutschen, Griechischen und Lateinischen*, 1(2. H), 163—179.
- Galperin, I. R. (1977). *Stylistics*. Moscow: Higher School Publishing House.
- Gnatchuk, H. (2015). A statistical analysis of English compounds in the newspaper style. *Mathematical Linguistics*, 1, 81—90.
- Gries, S. T., Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228—255.
- Gruodytė, R. (1994). Keletas administracinio stiliaus raiškos atvejų. *Kalbos kultūra [nuo 2014 m. – Bendrinė kalba]*, 66, 47—52.
- Grzybek, (2005). A Study on Russian Graphemes. (V. N. Toporov, Ed.) *Język - ličnost' - tekst. Sbornik statej k 70-letiju T.M. Nikolaevoj*, 237—263.
- Grzybek, (2007). History and methodology of word length studies. *Contributions to the Science of Text and Language*, 15—90. Springer.
- Grzybek, (2016). Word Length in Estonian Prose. *Trames*, 20(2), 145—175.
- Grzybek, P., Köhler, R. (2011). Viribus Quantitatis. *Exact methods in the study of language and text: dedicated to Gabriel Altmann on the occasion of his 75th birthday*, 62, v—xi. Walter de Gruyter.
- Grzybek, P., Rusko, M. (2009). Letter, Grapheme and (Allo-)Phone Frequencies: The Case of Slovak. *Glottology*, 2(1), 30—48.
- Grzybek, P., Stadlober, E., Kelih, E., Antić, G. (2005). Quantitative text typology: the impact of word length. *Classification—the Ubiquitous Challenge*, 53—64. Springer.

- Gu, Z., Eils, R., Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847—2849.
- Guerini, M., Pepe, A., Lepri, B. (2012). Do linguistic style and readability of scientific abstracts affect their virality? *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, 475—478.
- Hall, G. (2014). Pedagogical stylistics. *The Routledge handbook of stylistics*, 239—252.
- Halliday, M. A. (1971). Language in a social perspective. *Educational Review*, 23(3), 165—188.
- Halliday, M. A. (1994). Language as social semiotic. *Language and literacy in social practice* (23—43). Multilingual Matters Clevedon, England.
- Hausenblas, K. (1994). Approach to Standard Language, its Culture, and to Style. *The Prague School of Structural and Functional Linguistics: A Short Introduction*, 41, 311—332.
- Havránek, B. (2014 [1964]). The functional differentiation of the standard language. *Chapters from the history of Czech functional linguistics*, 27—40.
- Hernández-Campoy, J. M. (2016). *Sociolinguistic styles*. John Wiley ir Sons.
- Hoover, D. L. (2003). Another perspective on vocabulary richness. *Computers and the Humanities*, 37, 151—178.
- Hřebíček, L. (1986). Cohesion in Ottoman poetic texts. *Archiv orientální*, 54, 252—256.
- Hřebíček, L. (1996). Word frequency and word location in a text. *Archiv orientální*, 64(3), 339—347.
- Husson, F., Josse, J., Le, S., Mazet, J. (2018). Package ‘FactoMineR’. Package FactorMineR. Nuskaityta iš <https://cran.r-project.org/web/packages/FactoMineR/index.html>
- Yule, G. U. (1925). II. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FR S. *Philosophical transactions of the Royal Society of London. Series B, containing papers of a biological character*, 213.402-410, 21—87.
- Jakobson, R. (1960). Linguistics and poetics. *Style in language*, 350—377. MA: MIT Press.
- Janilionis, V., Morkevičius, V. ir Rauleckas, R. (2008). *Statistinių kiekybinių duomenų analizė su SPSS ir Stata*. Pavyzdinis metodologinis mokomasis studijų paketas. Kaunas: Kauno technologijos universitetas.
- Jolliffe, I. T. (2002). Principal components in regression analysis. *Principal component analysis*, 167—198.

- Juola, P., Mikros, G. K., Vinsick, S. (2018). Correlations and Potential Cross-Linguistic Indicators of Writing Style. *Journal of Quantitative Linguistics*, 26(2), 1—26.
- Karlgren, J., Cutting, D. (1994). Recognizing text genres with simple metrics using discriminant analysis. *Proceedings of the 15th conference on Computational linguistics*, 2, 1071—1075.
- Karttunen, L., Chanod, J.-P., Grefenstette, G., Schille, A. (1996). Regular expressions for language engineering. *Natural Language Engineering*, 2(4), 305—328.
- Kassambara, A. (2017). *Practical Guide To Principal Component Methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (T. 2). STHDA.
- Kelih, E., Altmann, G. (2015). A continuous model for polysemy. *Glottometrics*, 31, 31—37.
- Kelih, E., Rovenchak, A., Buk, S. (2014). Analysing h-point in lemmatised and nonlemmatised texts. *Studies in quantitative linguistics*, 17, 81—93.
- Kitchen, C. M. (2009). Nonparametric vs parametric tests of location in biomedical research. *American journal of ophthalmology*, 147, 571—572.
- Kniūkšta, (2003). Naujas kodeksas, senos kalbos problemos. *Bendrinė kalba (iki 2014 metų – Kalbos kultūra)*, 76, 66—79.
- Kniūkšta, (2005). Administracinė kalba ir jos vartoseną. *Gimtasis žodis*, 5, 2—3.
- Köhler, R. (1993). Synergetic linguistics. *Contributions to quantitative linguistics*, p. 41—51, Springer, Dordrecht.
- Köhler, R. (2003). Syntaktische Strukturen–Eigenschaften und Zusammenhänge. *LDV Forum*, 18(1/2), 299—320
- Köhler, R. (2005a). Gegenstand und Arbeitsweise der Quantitativen Linguistik (Aims and methods of Quantitative Linguistics). *Quantitative Linguistik/Quantitative Linguistics: Ein internationales Handbuch/An International Handbook*, 1—16.
- Köhler, R. (2005b). Synergetic Linguistics. *Quantitative Linguistik - Quantitative Linguistics. Ein internationales Handbuch*, 760—774. de Gruyter, Berlin/ New York.
- Köhler, R. (2005c). Properties of lexical units and systems (Eigenschaften lexikalischer Einheiten und Systeme). *Quantitative Linguistik / Quantitative Linguistics*, 305—311.
- Köhler, R. (2007). Quantitative analysis of syntactic structures in the framework of synergetic linguistics. *Aspects of Automatic Text Analysis*, 191—209. Springer, Berlin, Heidelberg.
- Köhler, R. (2012). *Quantitative syntax analysis* (T. Vol. 65). Walter de Gruyter.

- Köhler, R. (2015). Linguistic Motifs. *Sequences in language and text*, 69, 89—108.
- Köhler, R., Naumann, S. (2008). Quantitative text analysis using L-, F-and T-segments. *Data analysis, machine learning and applications*, 637—645. Springer, Berlin, Heidelberg.
- Köhler, R., Altmann, G. (2000). Probability Distributions of Syntactic Units and Properties. *Journal of Quantitative Linguistics*, 7(3), 189—200.
- Köhler, R., Naumann, S. (2016). Syntactic Text Characterisation Using Linguistic S-Motifs. *Glottometrics*, 34, 1—8.
- Kontutytė, E. (2007). Ką apima sąvoka “dalykinė kalba”? *Kalbotyra*, 57, 149—159.
- Korpela, J. (2001). *A tutorial on character code issues*.
- Kovalevskaitė, J. (2006). Dabartinės lietuvių kalbos tekstynas–10 metų kaupimo ir naudojimo patirtis. *Prace Baltystyczne*, 3, 231—241
- Koženiauskienė, R. (1999). *Retorika: iškalbos stilistika*. Mokslo ir enciklopedijų leidybos institutas.
- Koženiauskienė, R. (2011). Steigiamojo Seimo sveikinamųjų kalbų patosas. *Parlamento studijos. Mokslo darbai*, 10, 50—79.
- Koženiauskienė, R. (2013a). Retorika ir stilistika – visumos ir dalies santykis. *Šiuolaikinės stilistikos kryptys ir problemos*, 89—99.
- Koženiauskienė, R. (2013b). *Retorinė ir stilistinė publicistinių tekstų analizė*. Vilnius: Vilniaus universiteto leidykla.
- Kubát, M. (2015). *Kvantitativní analýza žánrů* [online]. Disertační práce, Univerzita Palackého v Olomouci, Filozofická fakulta, Olomouc. Nuskaityta iš <https://theses.cz/id/f3e0uv/>
- Kubát, M. (2016). *Kvantitativní analýza žánr* [Quantitative analysis of genres]. Ostrava: University of Ostrava.
- Kubát, M., Čech, R. (2016a). Quantitative Analysis of US Presidential Inaugural Addresses. *Glottometrics*, 34, 14—27.
- Kubát, M., Čech, R. (2016b). Text length and the thematic concentration of text. *Mathematical Linguistics*, 2, 5—13.
- Kubát, M., Matlach, V., Čech, R. (2014). *Studies in Quantitative Linguistics 18: QUITA-Quantitative Index Text Analyzer*. RAM-Verlag.
- Kubát, M., Milička, J. (2013). Vocabulary richness measure in genres. *Journal of Quantitative Linguistics*, 339—349.
- Lapinskas, R. (2003). *Įvadas į statistiką su R*. Vilnius: Vilniaus universiteto leidykla.
- Leonavičienė, A. (2004). Viešumo ir privatumo sankirtos: tiesioginė kalba politinėje spaudoje. *Darbai ir dienos*, 38, 149—166.

- Leonavičienė, A. (2005). Spaudos tekstų heterogeniškumas funkcinių stilių sandūros aspektu. *Kalbotyra*, 55, 38—46.
- Leonavičienė, A. (2007). Publicistinio ir šnekamojo stiliaus sandūra dabartinėje spaudoje. *Bendrinė kalba (iki 2014 metų–Kalbos kultūra)*, 80, 85—96.
- Leonavičienė, A. (2010). Sakinių ilgis--publicistinio ir šnekamojo stiliaus sandūros tekstuose požymis. *Kalbotyra*, 62, 95—107.
- Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.
- Mačutek, J., Mikros, G. K. (2015). Menzerath-Altmann law for word length motifs. *Sequences in Language and Text*, 69, 125—131.
- Mandelbrot, B. (1959). A note on a class of skew distribution functions: Analysis and critique of a paper by HA Simon. *Information and Control*, 2(1), 90—99.
- Mandelbrot, B. (1961). On the theory of word frequencies and on related Markovian models of discourse. *Structure of language and its mathematical aspects*, 19, 190—219.
- Mangiafico, S. S. (2016). *Summary and analysis of extension program evaluation in R*. Rutgers Cooperative Extension: New Brunswick, NJ, USA.
- Marcinkevičienė, R. (2000). Tekstynų lingvistika: teorija ir praktika. *Darbai ir dienos*, 24, 7—64.
- Marcinkevičienė, R. (2002). Akademinės recenzijos retorika. *Filologija*, 7, 22—26.
- Martinez, A. R. (2012). Part-of-speech tagging. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1), 107—113.
- Mathesius, V. (1983). Functional linguistics. *PRAGUIANA – Some Basic and Less Known Aspects of the Prague Linguistic School*, 121—142.
- McIntosh, R. (1967). An Index of Diversity and the Relation of Certain Concepts to Diversity. *Ecology*, 48(3), 392—404.
- Miangah, T. M., Altmann, G. (2016). The relationship between polysemy and word length in Persian. *Mathematical Linguistics*, 2(1), 35—42.
- Miangah, T. M., Rezai, M. J. (2016). Persian text ranking using lexical richness indicators. *Glottometrics*, 35, 6—15.
- Mukaka, M. M. (2012). A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24, 69—71.
- Muntigl, P., Gruber, H. (2005). Introduction: approaches to genre. *Folia Linguistica*, 39, 1—18.
- Nemcová, E., Altmann, G. (2008). The phoneme–grapheme relation in Slovak. *Analysis of Script. Properties of Characters and Writing Systems*, 77—85. Berlin: de Gruyter (in press).

- Pagojienė, D. (2006). Vertinimas ir jo raiška publicistinės esė tekstuose. *Kalbos vienetų semantika ir struktūra: mokslinių straipsnių rinkinys, parengtas pagal konferencijos pranešimus*, 137—144.
- Pagojienė, D. (2009). *Esė stilius*. Daktaro disertacija, Vytauto Didžiojo universitetas.
- Pečkuvienė, L. (2013). Administracinės kalbos žodynas: normos ir vartoseną. *Socialinių mokslų studijos*, 5(2), 525—539.
- Pelegrinova, K., Altmann, G. (2017). The study of adverbials in Czech. *Glottometrics*, 37, 34—53.
- Petrėnienė, O. (2003). Terminų aiškinimas mokslo populiarinamuosiuose tekstuose. *Terminologija*, 10, 42—54.
- Petrėnienė, O. (2011). *Mokslo populiarinamųjų tekstų kalbos ypatybės*. Vilniaus universitetas.
- Petrėnienė, O. (2013). Klausiamieji sakiniai mokslo populiarinamuosiuose tekstuose. *Šiuolaikinės stilistikos kryptys ir problemos*, 125—132.
- Petrenz, P. (2012). Cross-lingual genre classification. *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the ACL*, 11—21.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin ir review*, 21(5), 1112—1130.
- Pikčilingis J. (1971b). *Kas yra stilius*. Vilnius: Vaga.
- Pikčilingis J. (1971a). *Lietuvių kalbos stilistika 1*. Vilnius: Mintis.
- Pike, R., Thompson, K. (1993). Hello World. *USENIX Winter*, 43—50.
- Popescu, I. I., Altmann, G. (2014). The lambda structure of language levels. *Glottometrics*, 27, 54—88.
- Popescu, I. I., Best, K.-H., Altmann, G. (2014). Unified modeling of length in language. *Language*, 2.
- Popescu, I.-I. (2009). *Word frequency studies* (T. 64). Walter de Gruyter.
- Popescu, I.-I., Čech, R., ir Altmann, G. (2011a). Vocabulary richness in Slovak poetry. *Glottometrics*, 22, 62—72.
- Popescu, I.-I., Čech, R., ir Altmann, G. (2011b). *The lambda-structure of texts*. Ram-Verlag Lüdenscheid.
- Popescu, I.-I., ir Altmann, G. (2008). On the regularity of diversification in language. *Glottometrics*, 17, 94—108.
- Popescu, I.-I., ir Altmann, G. (2015). A simplified lambda indicator in text analysis. *Glottometrics*, 30, 19—44.
- Popescu, I.-I., Mačutek, J., ir Altmann, G. (2009). *Aspects of word frequencies*. RAM-Verlag Lüdenscheid.

- Popescu, I.-I., Mačutek, J., Kelih, E., Čech, R., Best, K.-H., ir Altmann, G. (2010). *Vectors and codes of text*. Lüdenscheid: RAM-Verlag.
- Popescu, I.-I., Miangah, T. M., Gnatchuk, H., Čech, R., Bodoc, A., Altmann, G. (2017). On Rank-Frequency Distributions in Poetry. *Glottometrics* 38, 30—54.
- Powers, D. M. (1998). Applications and explanations of Zipf's law. *Proceedings of the joint conferences on new methods in language processing and computational natural language learning*, 151—160. Association for Computational Linguistics.
- Ramage, D. (2007). Hidden Markov models fundamentals. *CS229 Section Notes*, 1.
- Rimkutė, E. (2006). Dabartinės lietuvių kalbos gramatinių formų vartoseną morfologiškai anotuotame tekste. *Lituanistica*, 34—55.
- Roelcke, T., Popescu, I. I., Altmann, G. (2017). Aspects of text concentration. *Glottometrics*, 36, 70—86.
- Rothschild, L. (1986). The distribution of English dictionary word lengths. *Journal of statistical planning and inference*, 14(2-3), 311—322.
- Schneider, W. (2006). XVI. Ernst Wilhelm Förstemann (1822-1906). *Glottometrics*, 12, 77—86.
- Shannon, C. E., Weaver, W. (1949). *The mathematical theory of communication*. Urbana: Univ. of Illinois Press.
- Sicilia-Garcia, E. I., Ming, J., Smith, F. J. (2003). Extension of Zipf's law to word and character n-grams for English and Chinese. *International Journal of Computational Linguistics & Chinese Language Processing*, 8(1), 77—102.
- Silagadze, Z. K. (1999). Citations and the Zipf-Mandelbrot's law. *arXiv preprint physics/9901035*.
- Smetonienė, I. (2000). Reklamos tekstų kompozicija. *Kalbotyra*, 48, 111—121.
- Smetonienė, I. (2001). Reklamos tekstų sintaksė. *Lituanistica*, 1, 80—92.
- Smetonienė, I. (2003). Tropai ir okaziniai dariniai reklamos tekstuose. *Bendrinė kalba (iki 2014 metų – Kalbos kultūra)*, 76, 126—130.
- Smetonienė, I. (2005). Sintaksinės figūros reklamos tekstuose. *Žmogus ir žodis*, 1, 102—105.
- Smetonienė, I. (2009). Naglio Šulijos orų prognozių stilius. *Žmogus ir žodis*, 11(1), 81—85.
- Smetonienė, I. (2013a). Buitinio stiliaus invazija į kitus funkcinis stilius. *Šiuolaikinės stilistikos kryptys ir problemos*, 150—158.
- Smetonienė, I. (2013b). Intertekstualumo formos reklamoje. *Bendrinė kalba (iki 2014 metų – Kalbos kultūra)*, 86, 167—178.

- Smetonienė, I. (2013c). Reklamos tekstų stilistinė skyryba. *Lietuvių kalba*, 7, 1—13.
- Smith, L. I. (2002). *A tutorial on principal components analysis* [interaktyvus]. Tech. rep., University of Otago.
- Snyman, D., Van Huyssteen, G. B., Daelemans, W. (2012). Cross-lingual genre classification for closely related languages. *23rd Annual Symposium of the Pattern Recognition Association of South Africa*, 132—137.
- Stamatatos, E., Fakotakis, N., Kokkinakis, G. (2000). Automatic text categorization in terms of genre and author. *Computational linguistics*, 26, 471—495.
- Steen, G. (1999). Genres of discourse and the definition of literature. *Discourse Processes*, 28, 109—120.
- Strauss, U., Altmann, G. (2003). Age and polysemy of words. *Glottometrics*, 6, 61—64.
- Strauss, U., Grzybek, P., Altmann, G. (2007). Word length and word frequency. *Contributions to the science of text and language*, 277—294. Springer, Dordrecht.
- Stukaitė, A. (2005). *Publicistinio stiliaus tekstų diferenciacija*. Magistro baigiamasis darbas, Vilniaus pedagoginis universitetas.
- Šlepikienė, R., Linkevičienė, N. (2014). Autoriaus ir skaitytojo santykių raiška publicistinio stiliaus tekstuose: redakcinių straipsnių kokybinė analizė. *Lietuvių kalba*, (7).
- Tamaoka, K., Altmann, G. (2005). Mathematical Modelling for Japanese Kanji Strokes in Relation to Frequency, Asymmetry and Readings. *Glottometrics*, 10, 16—29.
- Tatar, D., Lupea, M., ir Altmann, G. (2014). Hreb-like analysis of Eminescu's poems. *Glottometrics*, 28, 37—55.
- Těšitelová, M. (1992). *Quantitative linguistics*. John Benjamins.
- Tomczak, M., Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 21(1), 19—25.
- Tuzzi, A., Popescu, I.-I., Altmann, G. (2010). *Quantitative analysis of Italian texts*. RAM-Verlag Lüdenscheid.
- Uhlířová, L. (2005). Quantitative linguistics in the Czech Republic. *Quantitative Linguistik/Quantitative Linguistics: Ein internationales Handbuch/An International Handbook*, 129—135.
- Upton, G., ir Cook, I. (1996). *Understanding statistics*. Oxford University Press.
- Utka, A. (2004). *Statistinis tekstų funkcijų nustatymas*. Daktaro disertacija, Vytauto Didžiojo universitetas, Lietuvių kalbos institutas.

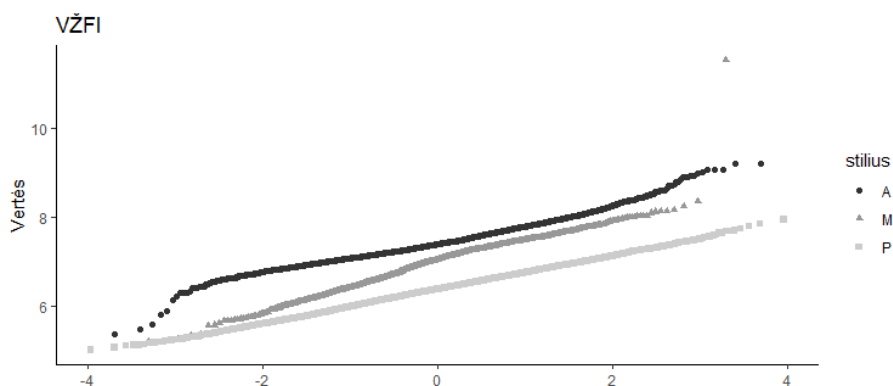
- Utka, A. (2005). Labai dažnų lietuvių kalbos žodžių ir žodžių formų ypatybės. *Lituanistica*, 1, 48—55.
- Vachek, J. (1962). On the interplay of external and internal factors in the development of language. *Lingua*, 11, 433—448.
- Vachek, J. (2014). On the functional hierarchy of spoken and written utterances. *Chapters from the history of Czech functional linguistics*, 78—94.
- van Oosten, P., Tanghe, D., Hoste, V. (2010). Towards an improved methodology for automated readability prediction. *7th Conference on International Language Resources and Evaluation (LREC 2010)*, 775—782.
- Vargha, A., Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, 25(2), 101—132.
- Versekėnaitė, A. (2008). Intertekstinės sekvencijos Dies iraeit XX a. kompozicijų sankirtos. *Musicology of Lithuania/Lietuvos muzikologija*, (9).
- Vladarskienė, R. (2000a). Apie funkcinis stilius ir registrus. *Acta Linguistica Lithuanica*, 42, 132—140.
- Vladarskienė, R. (2000b). Modalumas ir jo raiška valstybės dokumentuose. *Kalbos kultūra*, 90—95.
- Vladarskienė, R. (2002). Tarinio formos greta kelių veiksnių dokumentų kalboje. *Bendrinė kalba (iki 2014 metų – Kalbos kultūra)*, 75, 84—93.
- Vladarskienė, R. (2004). *Sintaksinių priemonių ypatybės valstybės dokumentų kalboje*. Liet. k. inst. I-kla.
- Vladarskienė, R. (2005). Lietuvių administracinės kalbos pokyčiai. *Žmogus ir žodis*, 7(1), 110—113.
- Vladarskienė, R. (2006). Keletas pastabų apie lietuvių administracinę kalbą. *Bendrinė kalba (iki 2014 metų–Kalbos kultūra)*, (79), 41—50.
- Vladarskienė, R. (2007). Lietuvių bendrinės ir administracinės kalbos santykis. *Bendrinė kalba (iki 2014 metų–Kalbos kultūra)*, (80), 55—63.
- Vladarskienė, R. (2012). Metaforiniai ekonomikos terminai. *Terminologija*, (19), 83—92.
- Vladarskienė, R. (2014). Prielinksnio dėl konstrukcijos administracinėje lietuvių kalboje. *Bendrinė kalba (iki 2014 metų – Kalbos kultūra)*, (87), 1—9.
- Vulanovic, R., Köhler R. (2005). Syntactic units and structures (Syntaktische Einheiten und Strukturen). *Quantitative Linguistik / Quantitative Linguistics*, 274—292
- Vulanović, R., Köhler, R. (2009). Word order, marking, and parts-of-speech systems. *Journal of Quantitative Linguistics*, 16(4), 289—306.
- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*. Springer.

- Wei, A., Liu, H. (2018). Typological Features of Zhuang from the Perspective of Word Frequency Distribution. *Glottometrics*, 44, 59—75.
- Westfall, H. (2014). Kurtosis as peakedness, 1905-2014. *RIThe American Statistician*, 68, 191—195.
- Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing*. Academic press.
- Wilson, A. (2009). Vocabulary richness and thematic concentration in internet fetish fantasies and literary short stories. *Glottology*, 2, 97—107.
- Wimmer, G., Altmann, G. (1995). A model of morphological productivity. *Journal of quantitative linguistics*, 2(3), 212—216.
- Wimmer, G., Altmann, G. (2007). Towards a unified derivation of some linguistic laws. *Contributions to the Science of Text and Language*, 329—337. Springer, Dordrecht.
- Xiao, W., Sun, S. (2018). Dynamic Lexical Features of PhD Theses across Disciplines: A Text Mining Approach. *Journal of Quantitative Linguistics*, 1—20.
- Zhang, C., ir Liu, H. (2015). A quantitative investigation of the genre development of modern Chinese novels. *Glottometrics*, 32, 9—20.
- Zörnig, P., Altmann, G. (2016). Activity in Italian presidential speeches. *Glottometrics*, 35, 38—48.
- Zörnig, P., Kelih, E., Fuks, L. (2016). Classification of Serbian texts based on lexical characteristics and multivariate statistical analysis. *Glottology*, 7, 41—66.
- Žilinskienė, V. (2001). Lietuvių ir latvių publicistikos leksikos ir morfologijos pagrindinės statistinės charakteristikos. *Lituanistica*, 69—79.
- Žilinskienė, V. (2002a). Gramatinių formų vartojimas lietuvių kalbos moksliniame stiliuje. *Acta Linguistica Lithuanica*, 46, 173—183.
- Žilinskienė, V. (2002b). Lietuvių kalbos administracinis stilius ir bendras jo leksikos vaizdas. *Viešoji politika ir administravimas*, 1, 51—54.
- Žilinskienė, V. (2002c). Žodžių formų vartojimas lietuvių kalbos dalykinio ir publicistinio stilių duomenimis. *Lituanistica*, (1), 106—117.
- Žilinskienė, V. (2003a). Administracinio lietuvių kalbos stiliaus skoliniai. *Viešoji politika ir administravimas*, (4), 74—81.
- Žilinskienė, V. (2003b). Gramatinių formų vartojimas lietuvių kalbos beletristiniame stiliuje. *Lituanistica*, 3, 75-84.
- Žilinskienė, V. (2005). Vardažodžių, įvardžių ir jų gramatinių formų vartojimas lietuvių kalbos stiliuose. *Lituanistica*, (4), 28—44.
- Žilinskienė, V. (2010). Skirtingų lietuvių kalbos stilių daiktavardžių kirčiuočių dažnumas. *Žmogus ir žodis*, 12(1), 111—121.
- Župerka, K. (1983). *Lietuvių kalbos stilistika*. Mokslas.

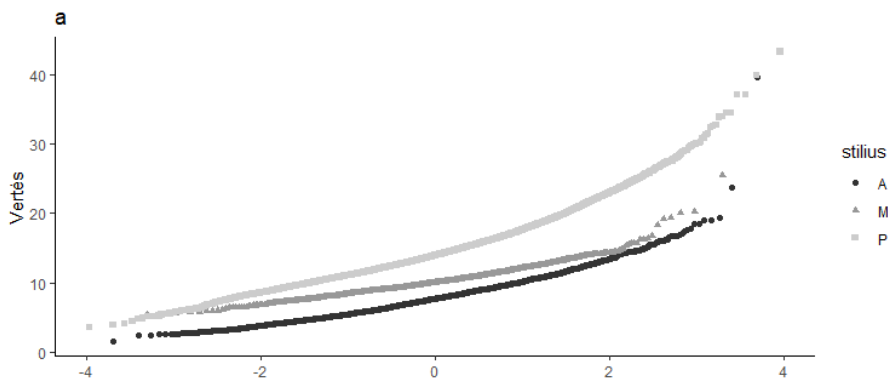
- Župerka, K. (1995). *Kalbos priemonių konkurencija kaip lietuvių kalbos stilistikos objektas*. Šiauliai: ŠPI.
- Župerka, K. (2000). Kalbos dabartiškumo nuorodos nelingvistiniuose tekstuose. *Acta linguistica Lithuanica*, 42, 124—131.
- Župerka, K. (2004). Reklamos teksto analizė: pragmatinės lingvistikos ir stilistikos sąveika. *Filologija*, (09), 55—59.
- Župerka, K. (2005). Bendrinė kalba ir funkciniai stiliai. *Kalbos kultūra*, 78, 56—66.
- Župerka, K. (2012). *Stilistika. III pataisytas ir papildytas leidimas*. Šiauliai: VšĮ Šiaulių universiteto leidykla.
- Župerka, K. (2016). Dabartinės viešosios kalbėsenos stiliaus polinkiai. *Acta humanitarica universitatis Saulensis*, 23, 171—184.

PRIEDAI

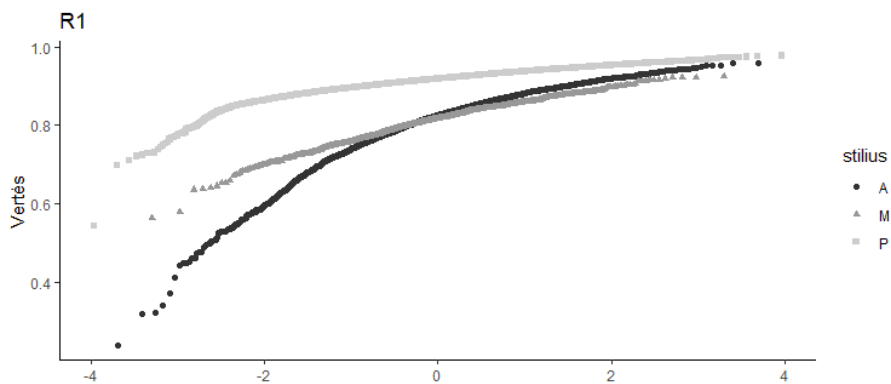
1 priedas. QQ (angl. *quantile-quantile*) diagramos (Field, Miles ir Field, 2012, 171-172), kuriose grafiškai iliustruoti administracinio, mokslinio ir publicistinio funkcinių stilių duomenų statistinis pasiskirstymas vizualiam stilių palyginimui pagal indikatorių raišką.



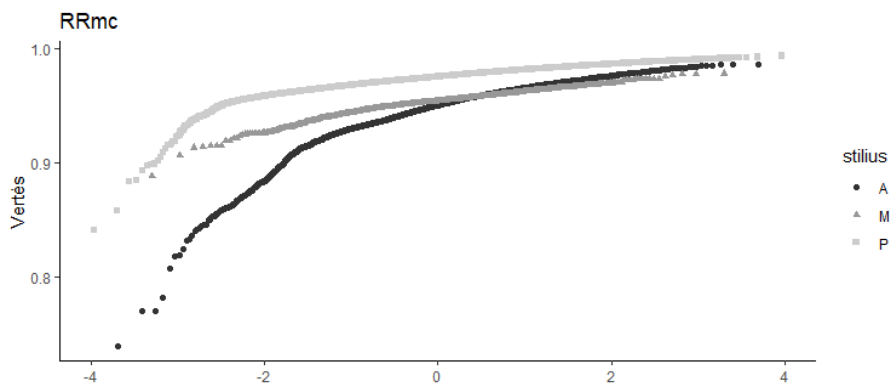
a. VŽFI verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



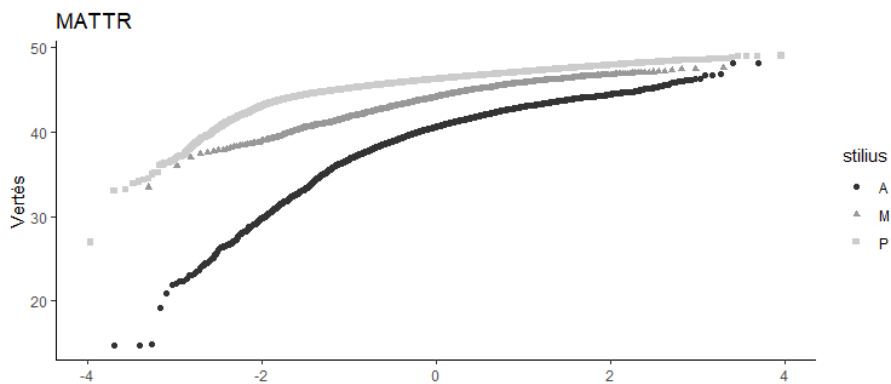
b. Indikatoriaus *a* verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



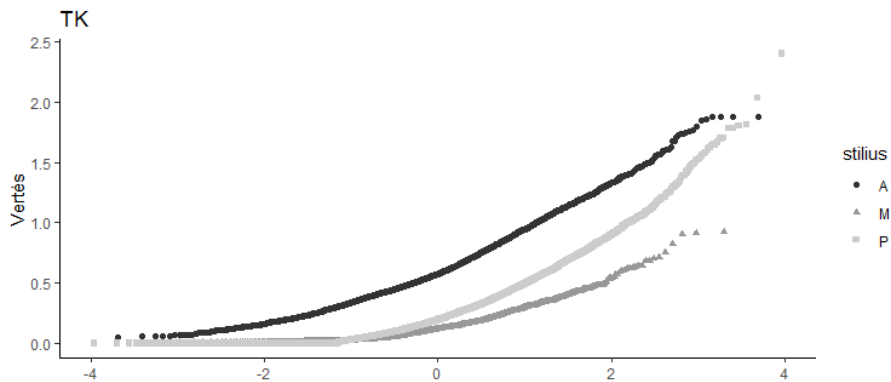
c. R_1 verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



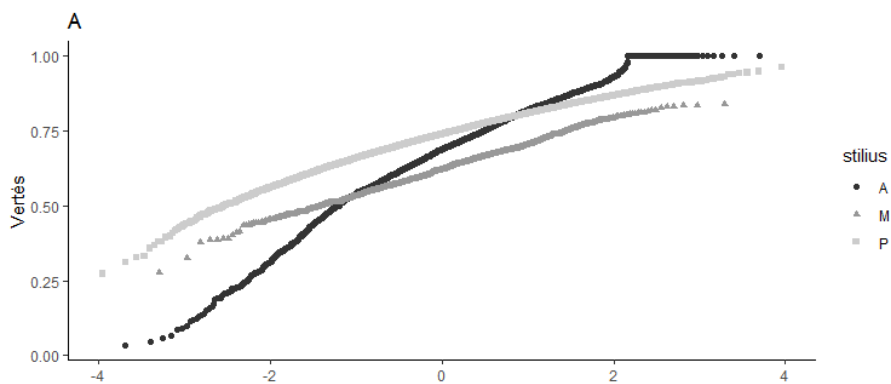
d. RR_{mc} verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



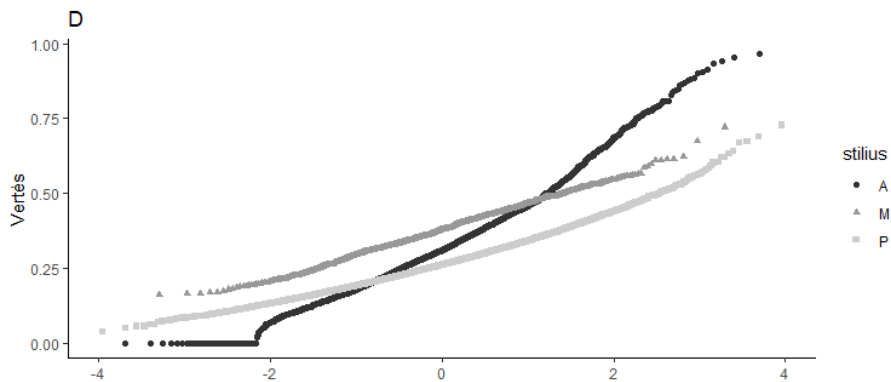
e. MATTR verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



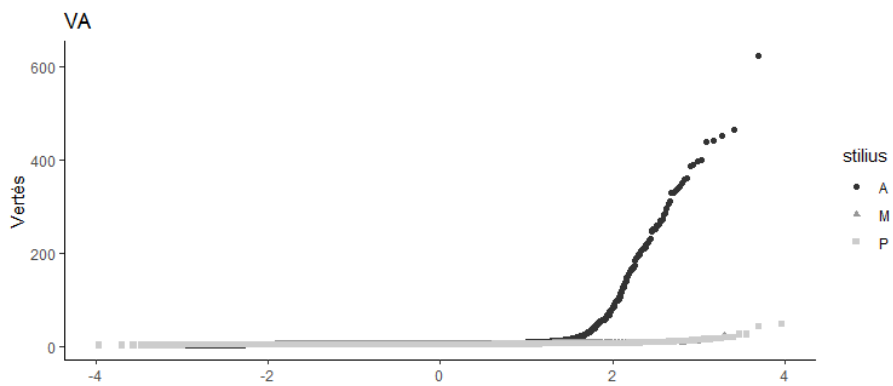
f. TK verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams



g. A verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams

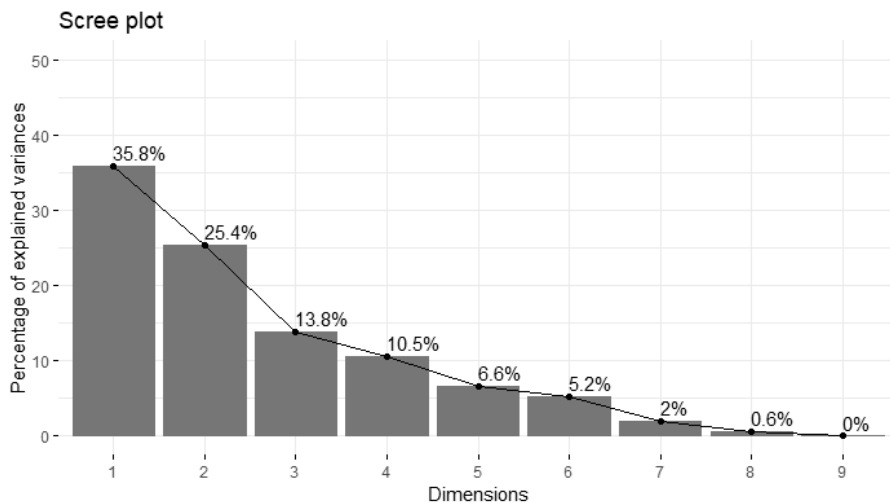


h. D verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams

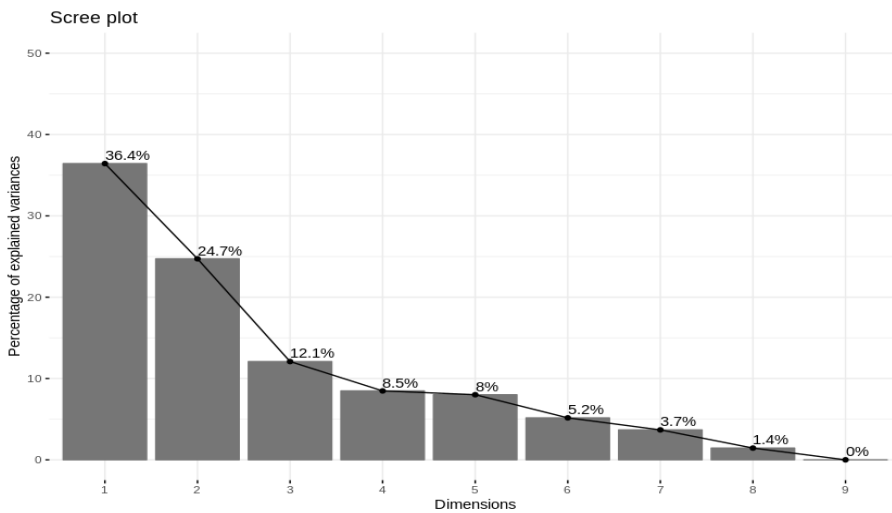


i. VA verčių statistinis pasiskirstymas administracinio (A), mokslinio (M) ir publicistinio (P) stiliaus tekstams

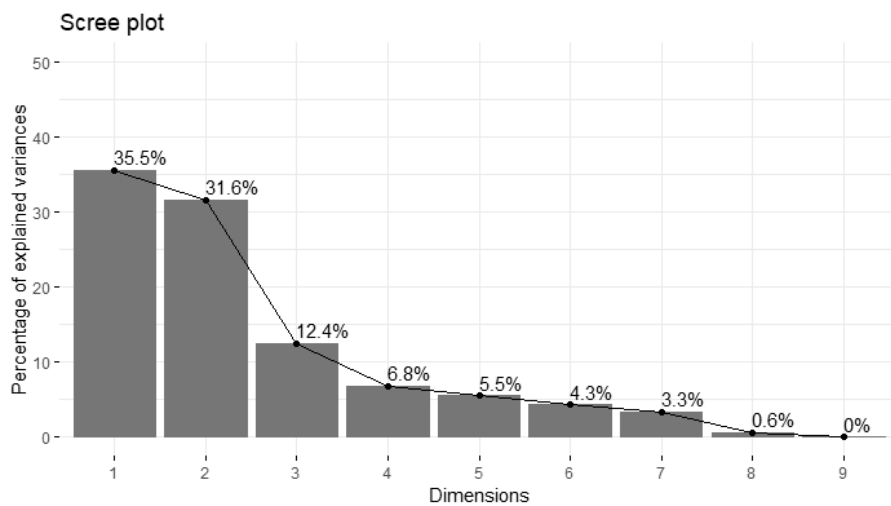
2 priedas. PKA: tikrinių reikšmių diagramos (angl. *scree plots*). Horizontalioje ašyje išdėstytos pagrindinės komponentės, vertikalioje – duomenų sklaidos procentas, kurį paaiškina atskiros komponentės.



a. PKA rezultatai: tikrinių reikšmių diagrama AST

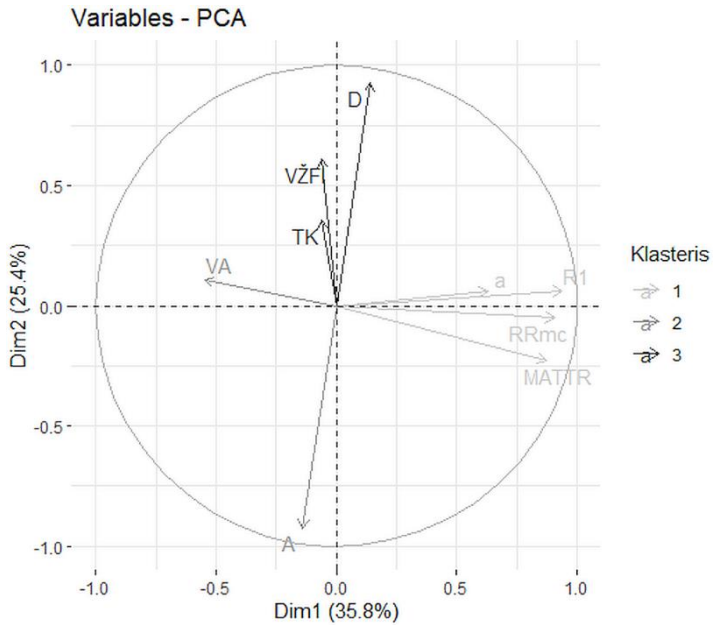


b. PKA rezultatai: tikrinių reikšmių diagrama MST

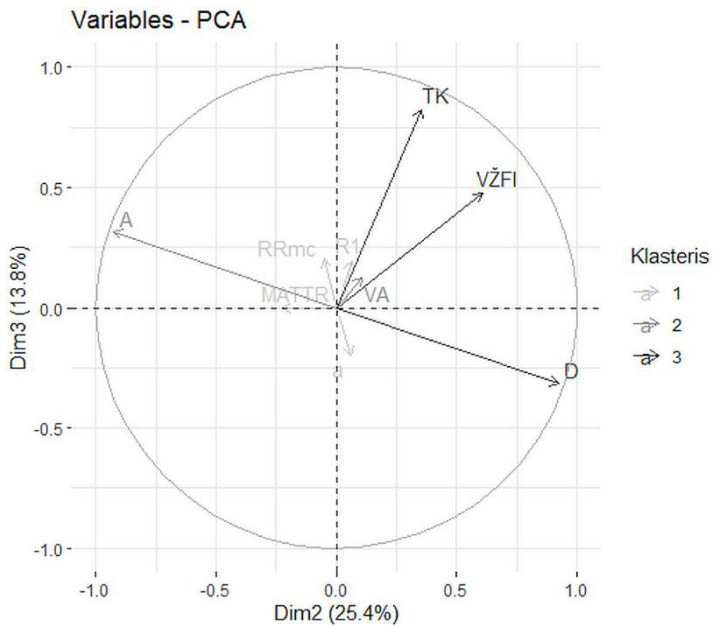


c. PKA rezultatai

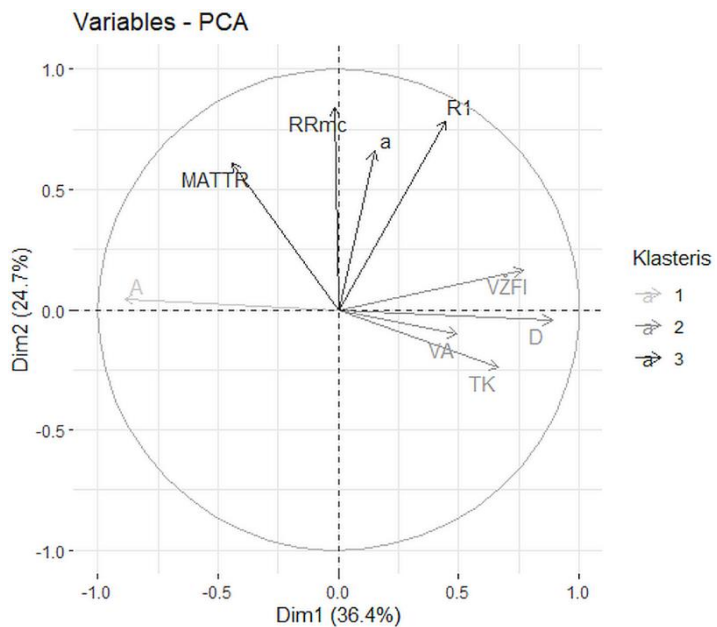
3 priedas. Indikatorių grupės pagal žymimų požymių panašumą (klasterinė analizė (Kassambara, 2017, 142)) atskirose pagrindinėse komponentėse AST, MST ir PST.



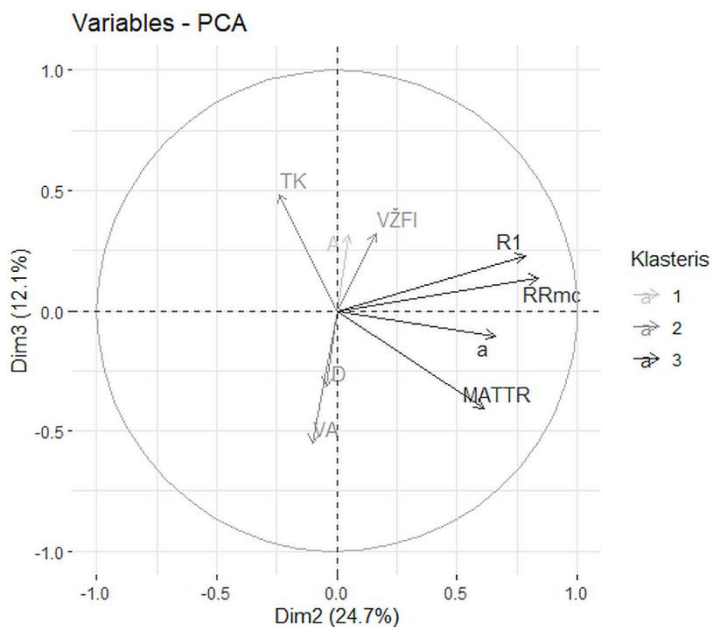
a. Indikatorių grupės pagal panašumą: AST (PK1 ir PK2)



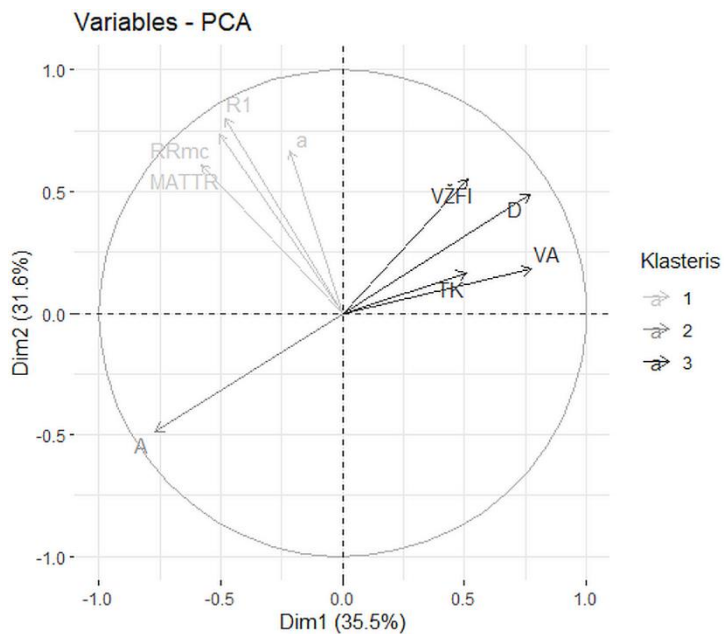
b. Indikatorių grupės pagal panašumą: AST (PK2 ir PK3)



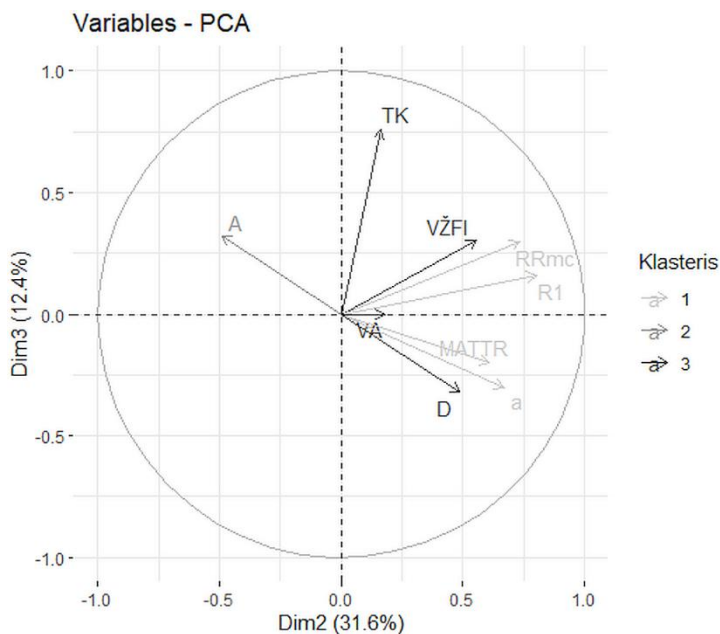
c. Indikatorių grupės pagal panašumą: MST (PK1 ir PK2)



d. Indikatorių grupės pagal panašumą: MST (PK2 ir PK3)



e. Indikatorių grupės pagal panašumą: PST (PK1 ir PK2)



f. Indikatorių grupės pagal panašumą: MST (PK2 ir PK3)

4 priedas. Dispersinė analizė: neparametrinių testų rezultatai, gauti panaudojus *R* bibliotekos **npvm** funkciją **nonpartest()**

Testo pavadinimas	Testo statistika	df1	df2	<i>p</i>-vertė	Permutacijos testo <i>p</i>-vertė
ANOVA tipo testas	3758,227	8,517	30246,44	0	0
McKeon'o aproksimuotas Lawley Hotelling testas	4173,918	16,000	31237,87	0	0
Muller'io aproksimuotas Bartlett-Nanda-Pillai testas	2287,823	16,003	38190,00	0	0
Wilks'o Lambda	3151,217	16,000	38184,00	0	0

5 priedas. Indikatorių verčių bazinės statistikos

Administracinis FS												
	n	trūkstamos	unikalios	vidurkis	Gini vid. skirtumas	,05	,10	,25	,50	,75	,90	,95
VŽFI	4527	0	4412	7,43	0,41	6,88	7,00	7,17	7,40	7,66	7,91	8,06
	Žemiausios vertės: 5,4; 5,5; 5,6; 5,8; 5,9; aukščiausios vertės: 9,1; 9,1; 9,1; 9,2; 9,2											
a	4527	0	3685	7,9	2,69	4,41	4,96	6,11	7,73	9,33	10,9	12,1
	Žemiausios vertės: 5,4; 5,5; 5,6; 5,8; 5,9; aukščiausios vertės: 9,1; 9,1; 9,1; 9,2; 9,2											
R ₁	4527	0	4379	0,81	0,08	0,66	0,71	0,77	0,83	0,87	0,89	0,91
	Žemiausios vertės: 0,24; 0,32; 0,32; 0,34; 0,37; aukščiausios vertės: 0,95; 0,95; 0,95; 0,96; 0,96											
RR _{mc}	4527	0	4339	0,95	0,02	0,91	0,92	0,94	0,95	0,96	0,97	0,97
	Žemiausios vertės: 0,74; 0,77; 0,77; 0,78; 0,81; aukščiausios vertės: 0,98; 0,99; 0,99; 0,99; 0,99											
MATTR	4527	0	4489	39,77	3,83	32,4	35,1	38,2	40,6	42,3	43,4	43,4
	Žemiausios vertės: 15; 15; 15; 19; 21; aukščiausios vertės: 47; 47; 47; 48; 48											
TK	4527	0	4224	0,63	0,34	0,21	0,28	0,41	0,58	0,82	1,06	1,19
	Žemiausios vertės: 0,05; 0,06; 0,06; 0,06; 0,06; aukščiausios vertės: 1,85; 1,86; 1,88; 1,88; 1,88											
A	4527	0	3138	0,67	0,16	0,40	0,49	0,59	0,69	0,78	0,85	0,89
	Žemiausios vertės: 0,03; 0,05; 0,06; 0,07; 0,09; aukščiausios vertės: 0,97; 0,97; 0,96; 0,98; 1,00											
D	4527	0	3138	0,33	0,16	0,11	0,15	0,22	0,31	0,41	0,51	0,60
	Žemiausios vertės: 5,4; 5,5; 5,6; 5,8; 5,9; aukščiausios vertės: 0,91; 0,93; 0,94; 0,95; 0,97											
VA	4527	0	3947	12,47	13,17	4,47	4,76	5,34	6,26	7,64	11,7	22,8
	Žemiausios vertės: 3,1; 3,4; 3,5; 3,5; 3,5; aukščiausios vertės: 437,3; 440,6; 450,7; 463,0; 622,2											
Mokslinis FS												
	n	trūkstamos	unikalios	vidurkis	Gini vid. skirtumas	,05	,10	,25	,50	,75	,90	,95
VŽFI	1025	0	1024	6,99	0,59	6,06	6,27	6,61	7,05	7,37	7,58	7,75
	Žemiausios vertės: 5,2; 5,3; 5,3; 5,4; 5,5; aukščiausios vertės: 8,1; 8,2; 8,3; 8,4; 11,5											
a	1025	0	1019	10,2	2,17	7,29	7,91	8,92	10,01	11,3	12,1	13,7
	Žemiausios vertės: 5,4; 5,7; 5,7; 5,9; 5,9; aukščiausios vertės: 19,2; 19,4; 20,0; 20,3; 25,4											
R ₁	1025	0	1021	0,81	0,06	0,72	0,74	0,77	0,82	0,85	0,87	0,88
	Žemiausios vertės: 0,56; 0,58; 0,64; 0,64; 0,64; aukščiausios vertės: 9,2; 9,2; 9,2; 9,2; 9,2											
RR _{mc}	1025	0	1009	0,95	0,01	0,93	0,94	0,95	0,95	0,96	0,96	0,97
	Žemiausios vertės: 0,89; 0,91; 0,91; 0,91; 0,91; aukščiausios vertės: 0,98; 0,98; 0,98; 0,98; 0,98											
MATTR	1025	0	1025	43,76	2,25	40,1	40,9	42,5	44,1	45,3	46,1	46,4
	Žemiausios vertės: 33 36 37 37 38; aukščiausios vertės: 47 47 47 47 48											
TK	1025	0	1025	0,16	0,15	0,01	0,02	0,04	0,12	0,23	0,35	0,44
	Žemiausios vertės: 0,00; 0,01; 0,01; 0,01; 0,01; aukščiausios vertės: 0,75; 0,83; 0,90; 0,92; 0,93											
A	1025	0	1012	0,62	0,098	0,48	0,51	0,56	0,62	0,68	0,74	0,77
	Žemiausios vertės: 0,28; 0,32; 0,38; 0,39; 0,39; aukščiausios vertės: 0,83; 0,83; 0,83; 0,84; 0,84											

D	1025	0	1012	0,38	0,098	0,23	0,26	0,32	0,38	0,44	0,49	0,52
	Žemiausios vertės: 0,16; 0,16; 0,17; 0,17; 0,17; aukščiausios vertės: 0,61; 0,61; 0,62; 0,68; 0,72											
VA	1025	0	1025	5,31	1,01	3,99	4,24	4,72	5,21	5,77	6,44	6,88
	Žemiausios vertės: 3,2; 3,3; 3,3; 3,4; 3,4; aukščiausios vertės: 8,9; 9,1; 9,1; 9,7; 22,5											
Publicistinis FS												
	n	trūkstamos	unikalios	vidurkis	Gini vid. skirtumas	,05	,10	,25	,50	,75	,90	,95
VŽFI	13450	0	12928	6,34	0,43	5,73	5,86	6,12	6,38	6,64	6,86	6,99
	Žemiausios vertės: 5,0; 5,1; 5,1; 5,1; 5,1; aukščiausios vertės: 7,7; 7,7; 7,8; 7,9; 7,9											
a	13450	0	6161	14,36	3,88	9,45	10,4	11,9	13,9	16,3	18,9	20,9
	Žemiausios vertės: 3,6; 4,0; 4,1; 4,5; 4,8; aukščiausios vertės: 34,4; 34,5; 37,0; 39,8; 43,2											
R₁	13450	0	11226	0,92	0,03	0,88	0,89	0,90	0,92	0,93	0,94	0,95
	Žemiausios vertės: 0,54; 0,70; 0,71; 0,72; 0,72; aukščiausios vertės: 0,97; 0,97; 0,97; 0,97; 0,98											
RR_{mc}	13450	0	10290	0,97	0,008	0,96	0,97	0,97	0,98	0,98	0,98	0,98
	Žemiausios vertės: 0,84; 0,86; 0,88; 0,88; 0,89; aukščiausios vertės: 0,99; 0,99; 0,99; 0,99; 0,99											
MATR	13450	0	13165	46,04	1,26	44,1	44,7	45,5	46,2	46,8	47,3	47,6
	Žemiausios vertės: 27 33 33 34 34; aukščiausios vertės: 49 49 49 49 49											
TK	13450	0	7479	0,26	0,26	0,00	0,00	0,07	0,19	0,38	0,59	0,74
	Žemiausios vertės: 0,00; 0,00; 0,00; 0,00; 0,00; aukščiausios vertės: 1,79; 1,80; 1,81; 2,03; 2,40											
A	13450	0	6387	0,73	0,09	0,60	0,63	0,69	0,74	0,79	0,83	0,85
	Žemiausios vertės: 0,27; 0,31; 0,33; 0,33; 0,36; aukščiausios vertės: 0,94; 0,94; 0,94; 0,95; 0,96											
D	13450	0	6387	0,27	0,09	0,15	0,17	0,21	0,26	0,31	0,37	0,40
	Žemiausios vertės: 0,04; 0,05; 0,06; 0,06; 0,06; aukščiausios vertės: 0,64; 0,67; 0,67; 0,69; 0,72											
VA	13450	0	9609	4,66	1,01	3,49	3,67	4,02	4,49	5,05	5,71	6,28
	Žemiausios vertės: 2,7; 2,7; 2,7; 2,8; 2,8; aukščiausios vertės: 19,9; 25,8; 25,9; 43,8; 47,3											

UŽRAŠAMS

UŽRAŠAMS

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p.: info@leidykla.vu.lt, www.leidykla.vu.lt
Tiražas 15 egz.