

LIETUVIŲ PUBLICISTIKOS TEKSTŲ KALBOS DALIŲ KORELIACINĖ IR KLASTERINĖ ANALIZĖ

VIDA ŽILINSKIENĖ

Apie tam tikrų kalbos dalių sintagminius santykius galima spręsti jau iš kalbos gramatinės sistemos aprašymo. Pavyzdžiui, yra žinoma, kad būdvardis morfologiškai ir sintaksiškai priklauso nuo daiktavardžio, kad viena iš įvardžio funkcijų ta, jog jis sakinyje gali pavaduoti daiktavardį, kad prievoksmis dažniausiai aiškina veiksmą, kad prielinksnis valdo tam tikrus daiktavardžio linksnius ir t. t.

Įvairiuose moksluose santykiai tarp atsitiktinių dydžių nustatomi koreliacinės analizės būdu. Koreliacinė analizė atskleidžia statistinius santykius tais atvejais, kai vienas atsitiktinis dydis sistemingai gauna tam tikrą reikšmę, didėjant arba mažėjant kitam dydžiui. Koreliacinė priklausomybė egzistuoja ir tada, kai vienas dydis susijęs su keliais kitais kintamaisiais dydžiais arba kai du dydžiai priklauso nuo kelių kitų dydžių [Смирнов, Дунин-Барковский, 1956; Kaminskienė, 1976].

Be abejonės, koreliacinę analizę galima taikyti ir kalbos dalių santykiams: ji gali padėti surasti vienos kalbos dalies aktyvumo (dažnumo didėjimo arba mažėjimo) sąryšį su kitos (arba kelių kitų) kalbos dalies aktyvumo kitimu.

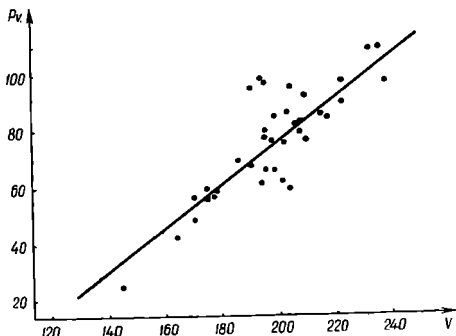
Kalbos dalių koreliacinių santykių tyrinėjimas ypač įdomus todėl, kad, patyrinęs ištraukas iš A. Gerceno, I. Turgenevo ir A. Gončiarovo kūrinių [Головин, 1974], buvo prieita prie išvadų, kad būdvardžių dažnumas nepriklausęs nuo daiktavardžių, kad nesą ryšio tarp daiktavardžių ir veiksmų, tarp daiktavardžių ir prielinksnių dažnumo A. Gerceno kalboje, o kitų autorių kalboje šie ryšiai pastebėti ir t. t. Atskirų autorių tekstų analizės rezultatų skirtumus bandoma aiškinti tuo, kad A. Gerceno kalba esanti publicistiškesnė ir racionalesnė. Vadinas, koreliaciniai santykiai turi gerokai priklausyti nuo kalbos stiliaus ir žanro. Tad įdomu pažiūrėti, kiek šie pastebėjimai tinka lietuvių publicistikai.

Šiame darbe nebandome išnagrinėti visus kalbos dalių koreliacinius santykius, o apsiribojame tik pačių koreliacinių santykių ir tų santykių krypties, pobūdžio ir laipsnio nustatymu¹.

¹ Koreliacinis santykis gali būti teigiamas arba neigiamas, tiesinis arba netiesinis. Teigiamas santykis būna tada, kai, didėjant vieno atsitiktinio dydžio dažnumui, kito dydžio dažnumas rodo aiškią didėjimo tendenciją; neigiamas — tada, kai, didėjant vieno dydžio dažnumui, kito dydžio dažnumas rodo panašią mažėjimo tendenciją. Koreliacinis santykis būna tam [tikro stiprumo, kurį parodo koreliacijos koeficiento skaitinė reikšmė.

Kalbos dalių koreliacinė analizė atlikta pagal lietuvių kalbos publicistikos dažnumų žodyno medžiagą. Kaip jau minėta [Žilinskienė, 1979, p. 79], sudarant šį žodyną, visa tyrinėjamoji medžiaga buvo suskirstyta penkiomis vienodo ilgio teminėmis grupėmis (po 60 000 pavartotų žodžių): A – vidaus ir užsienio politika, B – mokslas, technika, pramonė, žemės ūkis, C – švietimas, pedagogika, moralinės-etinės problemos, D – dailė, teatras, kinas, literatūros mokslas, E – sportas, skelbimai ir straipsniai, nepatekę į pirmąsias keturias grupes. Iš kiekvieno laikraščio ar žurnalo atsitiktinai imta viena 1000 žodžių praba.

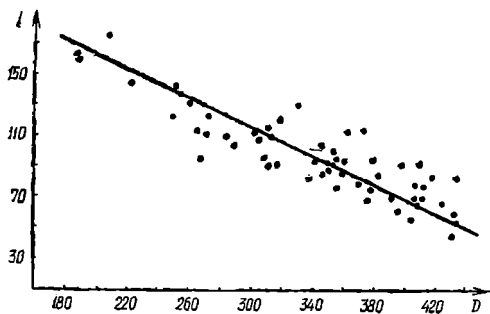
Kad nereikėtų iš karto imtis sudėtingų prabino koreliacijos koeficiento skaičiavimų tokiais atvejais, kai koreliacinio ryšio iš tikrųjų nėra, pradžioje buvo patikrinta prielaida apie galimą koreliacinio ryšio egzistavimą koreliacinių laukų metodu. Buvo paimtos dviejų tyrinėjamų kalbos dalių dažnumų sekos vienos teminės grupės prabose, ir šie dažnumai atitinkamu masteliu atidėti koordinacinių ašyse. Pvz., 1 brėžinyje pateikiami B temos veiksmažodžių irrieveiksmių, o 2 brėžinyje C temos daiktavardžių ir įvardžių santykiai. Gauti vadinamieji koreliaciniai laukai, arba taškinės diagramos. Kaip matyti iš brėžinių, abu koreliaciniai laukai ištempti pagal įstrižainę ir visi jų taškai išsidėstę apie tam tikrą tiesę. Tai aiškiai rodo, kad tarp tyrinėjamųjų kalbos dalių egzistuoja koreliaciniai santykiai². Tarpusavy brėžiniai skiriasi koreliacinio lauko pakrypimu koordinacinių atžvilgiu. 1 brėžinyje tu-



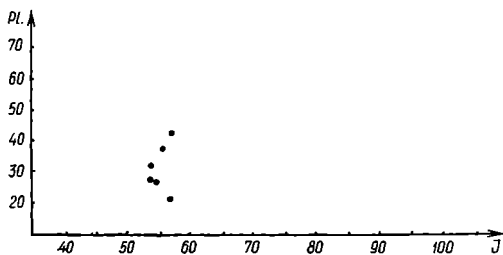
1 pav. B temos veiksmažodžio irrieveiksmio koreliacinis laukas

rime teigiamą koreliacinį ryšį: didėjant veiksmažodžių dažnumui, didėja irrieveiksmių dažnumas, ir, atvirkščiai, 2 brėžinys rodo neigiamą koreliacinį ryšį: didėjant daiktavardžių dažnumui, įvardžių dažnumas mažėja ir atvirkščiai.

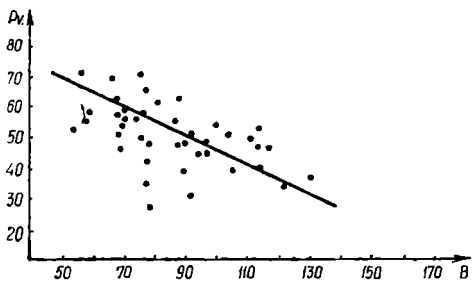
² Taškų susigrupavimas apie tiesę rodo dar ir tai, kad tarp šių kalbos dalių egzistuoja tiesinis koreliacinis ryšys, t. y. kad šių kalbos dalių dažnumai tiesiogiai priklauso vieni nuo kitų.



2 pav. C temos daiktavardžio ir įvardžio koreliacinis laukas



3 pav. D temos prielinksnio ir jungtuko koreliacinis laukas



4 pav. A temos būdvardžio irrieveksmio koreliacinis laukas

3 brėžinyje tokiu pat būdu pavaizduoti D temos prielinksnio ir jungtuko santykiai. Šiame koreliaciniame lauke taškai išsibarstę chaotiškai, – vadinasi, tarp šių kalbos dalių koreliacinio ryšio nėra. 4 brėžinyje pateiktas A temos būdvardžio ir prieveiksmio koreliacinis laukas. Nors čia taškai grupuojasi apie tiesę ne taip ryškiai, vis dėlto dar įmanoma nubrėžti tiesę, einančią iš kairiojo viršutinio brėžinio kampo į dešinįjį apatinį – tai galėtų rodyti, kad tarp šių kalbos dalių esama šio tūto silpno koreliacinio ryšio.

Koreliacinių laukų pobūdis leidžia manyti, kad tarp tam tikrų kalbos dalių egzistuoja skirtingo stiprumo koreliaciniai ryšiai ir kad tikslinga skaičiuoti prabinius koreliacinius koeficientus. Tų koeficientų skaitinės reikšmės turėtų konkrečiai parodyti ryšio stiprumą, o ženklas (pliusas arba minusas) – ryšio pobūdį. Iš tų pačių kalbos dalių koreliacijos koeficiento skaitinių reikšmių skirtingose teminėse grupėse būtų galima išsiaiškinti, ar tyrinėjamųjų tekstų tematika turi įtakos koreliaciniams santykiams. Pasirinktas prabinis koreliacijos koeficientas, kuris paprastai laikomas griežtesniu ir patikimesniu negu kiek lengviau nustatomas ranginis (Spirmeno) koreliacijos koeficientas [Головин, 1974, с. 6; Эдни, Морис, 1960, с. 167].

Pradiniai duomenys buvo bet kurių dviejų kalbos dalių sekos iš tam tikros teminės grupės prabų³, t. y. dvi atsitiktinių dydžių sekos x_1, x_2, \dots, x_n ir y_1, y_2, \dots, y_n . Apskaičiavus sekų aritmetinius vidurkius \bar{x}, \bar{y} ir empirines dispersijas σ_x, σ_y [Žilinskienė, 1979, p. 81], surandamas ryšio empirinis momentas (empirinė kovariacija)⁴:

$$m_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

³ Dažnumų sekos sudarytos taip. Ruošiant minėtą žodyną, buvo suskaičiuota, kiek kurių kalbos dalių yra kiekvienoje 1000 žodžių pavartojimų praboje. Kiekvienos teminės grupės prabos buvo išdėstytos tam tikra eile nuo pirmos iki šešiasdešimos. Eilė priimta sąlygiškai, bet vėliau jos griežtai laikytasi. Sudarant kiekvienos kalbos dalies dažnumų seką, imti tos kalbos dalies dažnumai pirmoje, antroje, ..., šešiasdešimtoje prabose. Pavyzdžiui, A teminės grupės būdvardžio dažnumų seka atrodo taip: 67, 40, 68, 76, 109, 82, 75, 93, 57, 119, 90, 105, 93, 133, 148, 107, 59, 116, 115, 98, 69, 113, 202, 69, 77, 124, 78, 192, 71, 98, 69, 95, 160, 71, 76, 70, 117, 90, 89, 78, 79, 78, 50, 86, 119, 150, 101, 58, 138, 116, 150, 121, 54, 101, 108, 79, 88, 128, 120, 88. Taigi kiekvienai kalbos daliai buvo sudarytos 5 dažnumų sekos, nes turėta penkios teminės grupės.

⁴ Visi skaičiavimai buvo atlikti universiteto Skaičiavimo centre elektronine skaičiavimo mašina „Minsk-32“ pagal S. Liberio sudarytą programą. Jeigu koreliacijos koeficientą tektų skaičiuoti „rankiniu būdu“, patogiau naudoti šiek tiek transformuotą koeficiento formulę [Гласс, Стэнли, 1976]:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$

Tada koreliacijos koeficientas skaičiuojamas pagal formulę:

$$r_{xy} = \frac{m_{xy}}{\sqrt{\sigma_x \sigma_y}} = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sigma_x \sigma_y}}$$

Kai dydžiai X ir Y būna pasiskirstę pagal normalųjį dėsnį (apie tai žr. [Žilinskienė, 1979]), koreliacijos koeficiento patikimumo intervalai surandami pagal formulę:

$$I = t_1 \frac{1 - r_{xy}^2}{\sqrt{n}},$$

kur t_1 – reikšmė randama Stjudento kriterijaus lentelėje [Митропольский, 1971] pagal pasirinktą reikšmingumo lygmenį. Visų teminių grupių kalbos dalių porų koreliacijos koeficientai kartu su patikimumo intervalais pateikiami koreliacinėse matricose (žr. 1–5 lent.).

Atliekant koreliacinę analizę, reikia nustatyti ryšio realumą, t. y. rasti, kurie koreliacijos koeficientai statistiškai reikšmingi ir kurie nereikšmingi (t. y. paaiškinami atsitiktiniais faktoriais)⁵. Koreliacinio ryšio realumas įvertinamas taip: jei koreliacijos koeficiento skaitinė reikšmė, įvertinus patikimumo intervalą, apima nulį, tai šie du dydžiai laikomi nekoreliuotais, jeigu neapima – koreliuoti⁶.

Kaip matyti iš lentelių, visų teminių grupių tekstuose daiktavardis yra reikšmingai neigiamai koreliuotas su įvardžiu, o reikšmingai teigiamai – su būdvardžiu. Tai reiškia, kad šiuose tekstuose daiktavardžio aktyvumas didėja įvardžio sąskaita ir kad daiktavardžio aktyvumo didėjimą lydi būdvardžio aktyvumo didėjimas ir atvirkščiai. Toks koreliacinis ryšys paaiškinamas tuo, kad sakinyje būdvardis dažniausiai eina šalia daiktavardžio kaip juo reiškiamos kalbos dalies pažymynys, o įvardis su daiktavardžiu dažniau būna susijęs ekskliuzyviniais ryšiais: dažniausi įvardžiai paprastai vartojami kaip daiktavardžio substitutai.

Daiktavardis reikšmingu neigiamu ryšiu yra susijęs su veiksmažodžiu, prieveiksmiu bei dalelyte: tyrinėtuose tekstuose daiktavardžio aktyvumas didėja veiksmažodžio, prieveiksmio, dalelytės sąskaita ir atvirkščiai. Toks daiktavardžio ir veiksmažodžio ryšys paaiškinamas tuo, kad sakinyje vienas veiksmažodis dažnai valdo kelis daiktavardžius, ir daiktavardžių skaičiaus padidėjimas sumažina veiksmažodžio užimamą santykinę dalį ir atvirkščiai. Galima, be to, manyti, kad vienų temų teks-

⁵ Teorinis koreliacijos koeficientas gali svyruoti nuo 0 (tarp atsitiktinių dydžių koreliacinio ryšio nėra) iki ± 1 (tarp dydžių yra funkcinė priklausomybė).

⁶ Vadovaujantis kitomis metodikomis, koreliacijos koeficiento reikšmingumas nustatomas paprasčiau (tai ypač aktualu, skaičiuojant „rankiniu būdu“): 1) esant dideliame sekos reikšmių skaičiui (t. y. kai $n > 50$), su 99% tikimybe galima tvirtinti, kad dydžiai X ir Y bus nekoreliuoti [Kaminskienė, 1976, p. 71], jeigu $r_{xy} < 2,58 \sqrt{\frac{1}{n-1}}$; 2) kritines koreliacijos koeficiento reikšmes galima surasti lentelėje [Liepa, 1974, p. 314], pagal 0,05 ar 0,01 reikšmingumo lygmenį.

	Daiktv.	Būdv.	Veiksm.	Priev.	Ivar.	Jungt.	Priel.	Dalel.
Daiktv.	-	0,49 ± 0,20	-0,61 ± 0,16	-0,48 ± 0,20	-0,55 ± 0,18	0,03 ± 0,26	-0,39 ± 0,22	-0,27 ± 0,24
Būdv.		-	-0,73 ± 0,12	-0,34 ± 0,23	-0,19 ± 0,25	0,32 ± 0,23	-0,50 ± 0,19	-0,32 ± 0,23
Veiksm.			-	0,47 ± 0,20	0,22 ± 0,25	-0,15 ± 0,25	0,26 ± 0,24	0,27 ± 0,24
Priev.				-	0,36 ± 0,23	0,05 ± 0,26	0,03 ± 0,26	0,28 ± 0,24
Ivar.					-	0,11 ± 0,26	0,12 ± 0,25	0,10 ± 0,26
Jungt.						-	-0,41 ± 0,22	-0,27 ± 0,24
Priel.							-	0,27 ± 0,24
Dalel.								-

B teminės grupės koreliacinė matrica

2 lentelė

	Daiktv.	Būdv.	Veiksm.	Priev.	Ivar.	Jungt.	Priel.	Dalel.
Daiktv.	-	0,53 ± 0,19	-0,72 ± 0,13	-0,71 ± 0,13	-0,75 ± 0,11	-0,12 ± 0,25	-0,32 ± 0,23	-0,68 ± 0,14
Būdv.		-	-0,45 ± 0,21	-0,48 ± 0,20	-0,55 ± 0,18	0,05 ± 0,26	-0,43 ± 0,21	-0,54 ± 0,18
Veiksm.			-	0,68 ± 0,14	0,57 ± 0,17	0,09 ± 0,26	0,14 ± 0,25	0,60 ± 0,16
Priev.				-	0,63 ± 0,16	0,10 ± 0,26	0,07 ± 0,26	0,69 ± 0,13
Ivar.					-	0,03 ± 0,26	0,20 ± 0,25	0,63 ± 0,16
Jungt.						-	0,25 ± 0,24	-0,09 ± 0,26
Priel.							-	0,18 ± 0,25
Dalel.								-

C teminės grupės koreliacinė matrica

3 lentelė

	Daiktv.	Būdv.	Veiksm.	Priev.	Īvar.	Jungt.	Priel.	Dalel.
Daiktv.	–	0,35 ± 0,23	–0,71 ± 0,13	–0,67 ± 0,14	–0,67 ± 0,14	–0,54 ± 0,18	–0,34 ± 0,23	–0,42 ± 0,21
Būdv.		–	–0,40 ± 0,22	–0,28 ± 0,24	–0,41 ± 0,22	–0,18 ± 0,25	–0,52 ± 0,19	–0,05 ± 0,26
Veiksm.			–	0,52 ± 0,19	0,66 ± 0,15	0,37 ± 0,22	0,15 ± 0,25	0,30 ± 0,23
Priev.				–	0,37 ± 0,22	0,39 ± 0,22	0,28 ± 0,24	0,33 ± 0,23
Īvar.					–	0,57 ± 0,18	0,19 ± 0,25	0,28 ± 0,24
Jungt.						–	0,07 ± 0,26	0,38 ± 0,22
Priel.							–	0,16 ± 0,25
Dalel.								–

5 lentelē

E teminés grupés korelācijas matrica

tams buvo būdingesnis veiksmažodinis stilius, kitų temų tekstams – daiktavardinis. Koreliacija tarp daiktavardžio irrieveiksmio kartoja daiktavardžio ryšį su veiksmažodžiu: sakinyjerieveiksmis dažniausiai aiškina veiksmažodį ir todėl jo dažnumas priklauso nuo veiksmažodžio. Daiktavardžio ir jungtuko koreliacinis ryšys įvairių teminių grupių tekstuose yra labai skirtingas. Trys koeficientai rodo, kad ryšio nėra. Daiktavardis yra reikšmingai neigiamai koreliuotas su prielinksniu visų grupių tekstuose, išskyrus B grupę. Viena gana netikėto fakto priežastimi, matyt, laikytina tai, kad tekste prielinksnių skaičius priklauso nuo tam tikrų veiksmažodžio formų skaičiaus. Kadangi kai kurie veiksmažodžiai reikalauja atitinkamų prielinksnių konstrukcijų ir, didėjant daiktavardžių skaičiui, prielinksnių skaičius, matyt, nedidėja – tuo pačiu darosi reliatyviai mažesnė ta teksto dalis, kurią užima prielinksnis, plg.: *Pokalbyje pranešėjas užsiminė apie dėstytojus. Pokalbio metu pranešėjas užsiminė apie dėstytojų darbą.*

Visose teminėse grupėse būvdvardis yra reikšmingai neigiamai koreliuotas su veiksmažodžiu, o neigiama jo koreliacija surieveiksmiu ir įvardžiu svyruoja atskirose grupėse. Geras nekoreliuotų reiškinų pavyzdys yra būvdvardžio ir jungtuko aktyvumas: čia visų koreliacijos koeficientų reikšmės artimos nuliui (išskyrus A temą), o koeficiento ženklas įvairuoja – jis vienur teigiamas, kitur neigiamas. Koreliacija tarp būvdvardžio ir prielinksnio primena koreliaciją tarp daiktavardžio ir prielinksnio (išskyrus E temą, kurioje šie dydžiai nekoreliuoti). Būvdvardžio ir dalelytės koreliacija skirtinga atskirose teminėse grupėse. Tai rodo, kad būvdvardžio ir kitų kalbos dalių koreliaciniai ryšiai tartum kartoja daiktavardžio santykius su kitomis kalbos dalimis.

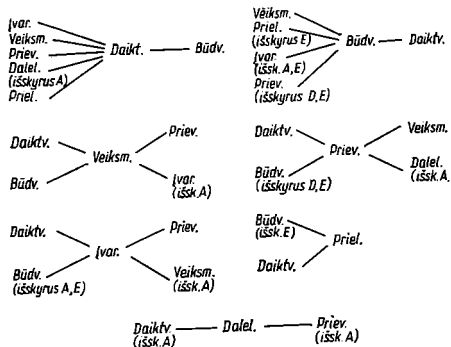
Veiksmažodžio irrieveiksmio dažnumai visur koreliuoti reikšmingai teigiamai: veiksmažodžių aktyvumo didėjimą lydirieveiksmių aktyvumo didėjimas ir atvirkščiai. Tokie pat koreliaciniai ryšiai jungia veiksmažodį su įvardžiu (išskyrus A temos tekstus, kuriuose šie dydžiai yra nekoreliuoti). Veiksmažodis su jungtuku nekoreliuotas: tik du koeficientai rodo silpną ryšį; kai kuriose grupėse svyruoja koeficiento ženklas. Varijuoja koreliaciniai santykiai tarp veiksmažodžio ir prielinksnio bei veiksmažodžio ir dalelytės dažnumų.

Prieveiksmio koreliacija su įvardžiu ir jungtuku visai artima veiksmažodžio koreliacijai su tomis kalbos dalimis. Prieveiksmio ir prielinksnio dažnumai nekoreliuoti – silpnas ryšys pastebimas tik C grupėje. Reikšmingas teigiamas koreliacinis ryšys yra tarprieveiksmio ir dalelytės, išskyrus A grupės tekstus.

Labai nepastovi įvardžio koreliacija su tarnybinėmis kalbos dalimis. Daugečiu atvejų koeficientai rodytų ryšio nebuvimą, bet vietomis tas ryšys būna reikšmingai teigiamas.

Nėra koreliacinio ryšio tarp jungtuko ir prielinksnio, jungtuko ir dalelytės, prielinksnio ir dalelytės dažnumų.

Apibendrinus aptartuosius koreliacijos koeficientus, matyti, kad reikšmingi koreliaciniai ryšiai egzistuoja tarp tokių kalbos dalių (į dešinę pusę atidėti teigiami ryšiai, į kairę – neigiami; kuo ryšys yra stipresnis, tuo ta kalbos dalis atidėta aukščiau; žr. 5 pav.):



5 pav. Visų kalbos dalių koreliaciniai santykiai

Taigi beveik visų teminių grupių tekstuose nominatyvinės kalbos dalys yra tarpusavy koreliuotos reikšmingais ryšiais ir tik kai kurios iš jų susijusios su tam tikromis tarnybinėmis kalbos dalimis. Tarnybinės kalbos dalys tarpusavy nekoreliuotos; jungtukas nekoreliuotas nė su viena kalbos dalimi⁷.

Apsvarsčius šiuos rezultatus, kyla mintis, kad tarpusavio aktyvumo atžvilgiu visas kalbos dalis galima suskirstyti į dvi grupes: vieną sudarytų daiktavardis ir būdvardis, kitą – veiksmazodis,rieveiksmis, įvardis. Kiekvienos grupės viduje kalbos dalys būtų koreliuotos teigiamai, o tarp skirtingoms grupėms priklausančių kalbos dalių egzistuočių neigiamas ryšys.

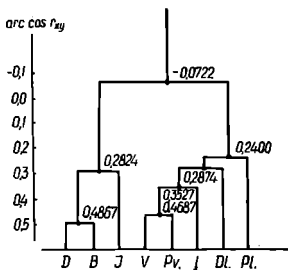
Šią prielaidą galima patikrinti klasterinės analizės (arba vadinamojo hierarchinio grupavimo⁸) būdu, leidžiančiu išryškinti labiausiai susijusių dydžių (šiuo atveju kalbos dalių) hierarchines grupes [Johnson, 1967; Levelt, 1970; Айвазян, Бежаева, Староверов, 1974].

⁷ Taip pat iš šių brėžinių ir koreliacinių matricų matyti, kad A ir E teminių grupių (A – vidaus ir užsienio politika, E – sportas, skelbimai ir įvairūs straipsniai) rezultatai yra labiausiai išsibarstę, o B ir C (B – mokslas, technika, pramonė, žemės ūkis, C – švietimas, pedagogika, moralinės-ėtinės problemos) – mažiausiai.

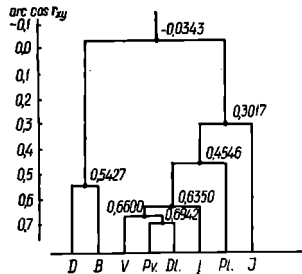
⁸ Hierarchinio grupavimo metodiką yra naudojęs A. Pakerys, nustatinėdamas lietuvių bendrinės kalbos balsių psichoakustinę hierarchiją [Pakerys, 1975] bei A. Girdenis ir A. Rosinas, klasifikuodami lietuvių bendrinės kalbos ir žemaičių tarmės daiktavardžių linksniaavimo tipus [Girdenis, Rosinas, 1977]. Minėtame A. Pakerio straipsnyje smulkiai aprašoma ir pati hierarchinio grupavimo metodika.

Žingsnis	A teminė grupė			B teminė grupė			C teminė
	kalbos dalis	tarpgrupinis atstumas r	$\cos r$	kalbos dalis	tarpgrupinis atstumas r	$\cos r$	kalbos dalis
1	D + B	1,0625	0,4867	Pv. + DI.	0,8025	0,6942	D + B
2	V + Pv.	1,0830	0,4687	V + Pv. + DI.	0,8500	0,6600	V + I
3	V + Pv + I	1,2103	0,3527	V + Pv. + DI. + I	0,8828	0,6350	V + I + Pv.
4	V + Pv. + I + DI.	1,2793	0,2874	D + B	0,9971	0,5427	V + I + Pv. + DI.
5	D + B + J	1,2845	0,2824	V + Pv. + DI. + I + PI.	1,0988	0,4546	V + I + Pv. + DI. + PI.
6	V + Pv. + I + DI. + PI.	1,3284	0,2400	V + Pv. + DI. + I + PI. + J	1,2643	0,3017	V + I + Pv. + DI. + PI. + J
7	D + B + J + V + Pv + I + DI. + PI.	1,6430	-0,0722	D + B + V + Pv. + DI. + I + PI. + J	1,6051	-0,0343	D + B + V + I + Pv. + DI. + PI. + J

Sutartiniai žymėjimai: D – daiktavardis, B – būdvardis, V – veiksmažodis, Pv. – prieveiks-



6 pav. A temos dendrograma



7 pav. B temos dendrograma

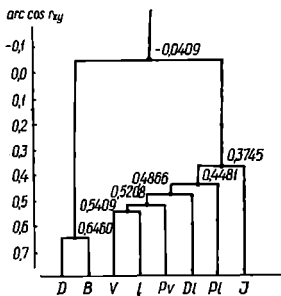
Išėities poziciją sudarė porinių koreliacijos koeficientų matricos. Kadangi kai kurie koreliacijos koeficientai yra neigiami, pirmiausia atliekama koreliacijos koeficiento transformacija $\text{arc cos } r$, paverčianti visus dydžius teigiamais⁹. Taigi pradžioje turime 8×8 eilių simetrinę matricą kiekvienai teminei grupei. Joje suieškome mažiausią reikšmę, siejančią dvi labiausiai susijusias kalbos dalis. Šią

⁹ Klasterinė analizė irgi atlikta ESM „Minsk-32“ pagal atitinkamą programą.

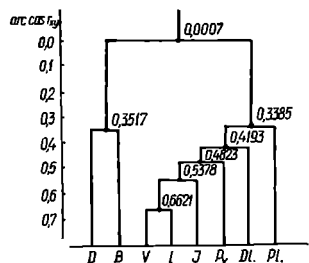
dendrogramoms

grupė		D teminė grupė			E teminė grupė		
tarpgrupinis atstumas r	$\cos r$	kalbos dalis	tarpgrupinis atstumas r	$\cos r$	kalbos dalis	tarpgrupinis atstumas r	$\cos r$
0,8685	0,6460	V + I	0,8472	0,6621	Pv. + DI.	1,0605	0,4884
0,9993	0,5409	V + I + J	1,0029	0,5378	I + PI.	1,0771	0,4739
1,0231	0,5208	V + I + J + Pv.	1,0675	0,4823	V + I + PI.	1,1555	0,4035
1,0626	0,4866	V + I + J + Pv. + DI.	1,1382	0,4193	D + B	1,1630	0,3966
1,1062	0,4481	D + B	1,2114	0,3517	Pv. + DI. + J	1,2334	0,3310
1,1870	0,3745	V + I + J + Pv. + DI. + PI.	1,2255	0,3385	V + I + PI. + Pv. + DI. + J	1,2984	0,2691
1,6117	-0,0409	D + B + V + I + J + Pv. + DI. + PI.	1,5701	0,0007	D + B + V + I + PI. + Pv. + DI. + J	1,5659	0,0049

mis, I – įvardis, J – jungtukas, PI. – prielinksnis, DI. – dalelytė



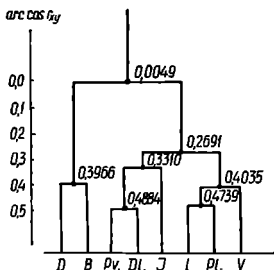
8 pav. C temos dendrograma



9 pav. D temos dendrograma

reikšmę įsimename, o tas dvi grupės sujungiamė į vieną, tada lieka 7×7 eilių matrica¹⁰. Po kiekvieno tolesnio grupavimo grupių skaičius sumažėja vienetu, nes į vieną sujungiamos dvi artimiausios grupės. Atlikti 7 grupavimai, kurių rezultatai pateikti 6 lentelėje. Pagal tą lentelę nubraižytos dendrogramos (6–10 brėž.).

¹⁰ Smulkiau apie tai žr. minėtame Pakerio straipsnyje.



10 pav. E temos dendrograma

Dendrogramos įtikinamai rodo, kad: a) visų temų tekstuose į vieną reliatyviai savarankišką grupę jungiasi daiktavardis su būdvardžiu, o į kitą — veiksmazodis, prieveiksmis, įvardis, b) labai nepastovii tarnybinių kalbos dalių vieta liudija, kad šiame darbe priimtas visos tyrinėjamos medžiagos skirstymas į nurodytas temines grupes yra nerelevantinis tarnybinių kalbos dalių atžvilgiu.

КОРРЕЛЯЦИОННЫЙ И КЛАСТЕРНЫЙ АНАЛИЗ ЧАСТЕЙ РЕЧИ ЛИТОВСКОЙ ПУБЛИЦИСТИКИ

Резюме

Цель данной работы — установление корреляционных связей между определенными частями речи в литовской публицистике.

Корреляционный и кластерный анализ проводился по материалу частотного словаря литовской публицистики. Описана методика подготовки исходных данных и построены корреляционные поля (точечные диаграммы) для предварительной проверки наличия корреляционных связей между некоторыми частями речи. Приведена методика расчета выборочного корреляционного коэффициента. Анализ указанных коэффициентов показал, что в публицистических текстах именные части речи в основном имеют значительную взаимную обусловленность (положительную или отрицательную), и лишь небольшое количество их связано с некоторыми служебными частями речи. Служебные части речи не коррелированы между собой. Союз не коррелирован ни с одной частью речи. На базе выборочных коэффициентов корреляции произведен кластерный анализ, выявляющий связи внутри набора частот частей речи, и получены дендрограммы для каждого подъязыка исследуемого материала. Анализ дендрограмм убедительно показал следующее: 1) во всех подъязыках в одну группу объединяются существительное и прилагательное, а в другую — глагол, наречие, местоимение; 2) отмечается большой разброс в расположении служебных частей речи.

LITERATŪRA

- Girdenis, Rosinas, 1977 – Girdenis A., Rosinas A. Lietuvių kalbos daiktavardžio linksniavimo tipų hierarchija ir pagrindinės jos kitimo tendencijos (bendrinės kalbos ir žemaičių tarmės duomenimis). – *Baltistica*, 1974, t. 13(2), p. 338–348.
- Johnson, 1967 – Johnson S. C. Hierarchical Clustering Schemes. – *Psychometrica*, 1976, vol. 32, No 2, p. 241–254.
- Kaminskienė, 1976 – Kaminskienė B. Matematinės statistikos elementai. II dalis. Koreliacinė-regresinė analizė. – V., 1976.
- Levelt, 1970 – Levelt W. J. M. Hierarchical Clustering Algorithms in the Psychology of Grammar. – In: *Advances in Psycholinguistics*. Amsterdam—London, 1970, p. 101–108.
- Liepa, 1974 – Liepa J. Biometrija. – Rīga, 1974.
- Pakerys, 1975 – Pakerys A. Lietuvių bendrinės kalbos balsių diferencinių požymių hierarchija. – *Kalbotyra*, 1975, t. 26(1), p. 37–47.
- Žilinskienė, 1979 – Žilinskienė V. Kalbos dalių statistinė charakteristika lietuvių publicistikoje. – *Kalbotyra*, 1979, t. 30(1), p. 78–88.
- Айвазян, Бежаева, Староверов, 1974 – Айвазян С. А., Бежаева З. А., Староверов О. В. Классификация многомерных наблюдений. – М., 1974.
- Гласс, Стэнли, 1976 – Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М., 1976.
- Головин, 1966 – Головин Б. Н. Опыт применения корреляционного анализа в изучении речи. – В кн.: *Вопросы статистической стилистики*. Киев, 1974, с. 5–16.
- Митропольский, 1971 – Митропольский А. К. Техника статистических вычислений. – М., 1971.
- Смирнов, Дунин-Барковский, 1965 – Смирнов Н. В., Дунин-Барковский И. В. Курс теории вероятностей и математической статистики для технических приложений. – М., 1965.
- Эдни, Морис, 1960 – Эдни Юл. Дж., Морис К. Дж. Теория статистики. – М., 1960.

Vilniaus V. Kapsuko universitetas
Eksperimentinės fonetikos laboratorija

Įteikta
1979 m. lapkričio 23 d.