

AGNĖ PUKEVIČIŪTĖ
Vytauto Didžiojo universitetas

Mokslinių tyrimų kryptys: sakininės lietuvių kalbos ypatybės.

ASTA KAZLAUSKIENĖ
Vytauto Didžiojo universitetas

Mokslinių tyrimų kryptys: sakininės lietuvių kalbos tyrimai: akustinės garsų ypatybės, fonotaktika, kirčiavimo dėsningumai, normos ir vartoseną, kalbos ritmika; taikomoji kalbotyra.

KLASTERINĖ TIPOLOGINIŲ KALBOS RITMO GRUPIŲ ANALIZĖ¹

Cluster Analysis of Typological Groups
of Language Rhythm

ANOTACIJA

Šiuo straipsniu siekiama atsakyti į du klausimus: 1) ar klasterinė analizė paranki ieškant ritmo tipologinių panašumų ir ja remiantis galima klasifikuoti iki šiol neanalizuotas kalbas ir 2) į kurį klasterį – skiemeninio ar kirtinio ritmo kalbų – patenka lietuvių ir latvių kalbos.

Tyrimui penki gimtakalbiai (latvis, lietuvis, švedas, islandė ir kinė) išvertė ir įrašė tą patį apie 1 min. trukmės tekstą. Programa PRAAT įrašai buvo anotuoti, naudojant Paolo Mairano

¹ Tyrimą finansavo Lietuvos mokslo taryba (sutarties Nr. LIT-5-4). Pradiniai jo rezultatai apibendrinti Agnės Pukevičiūtės magistro baigiamajame darbe (*Akustiniai lietuvių ir latvių kalbų ritmo koreliatai*, 2013, vadovė prof. dr. Asta Kazlauskienė). Straipsnio autorės nuoširdžiai dėkoja informantams, įrašiusiems ir padėjusiems segmentuoti tiriamąją medžiagą: prof. habil. dr. Pėterui Vanagui, Valgerður Bjarnadóttir, Magnus Liw, Huashan Li ir doc. dr. Aaušrai Saudargienei už pagalbą rengiant medžiagą statistinei analizei ir patarimus dirbant su SPSS programa.

ESMINIAI ŽODŽIAI: ritmas, kirtinio ritmo kalbos, skiemeninio ritmo kalbos, akustiniai kalbos ritmo koreliatai, klasterinė analizė, lietuvių, latvių, švedų, islandų, kinų kalbos.

KEYWORDS: Rhythm, Stress-timed languages, Syllable-timed languages, Acoustic correlates of rhythm, Cluster analysis, Lithuanian, Latvian, Swedish, Icelandic, Chinese.

sukurtą programą *Correlatore* (2.1 versija) buvo apskaičiuoti akustiniai ritmo koreliatai. Klasterinei analizei naudoti ir šios programos autoriaus Mairano (2011) gauti kitų kalbų ritmo tyrimo rezultatai. Duomenys suvesti į SPSS programą ir atlikta jų klasterinė analizė.

Atlikto tyrimo rezultatai rodo, kad hierarchinė ir dviejų žingsnių klasterinė analizė viena-reikšmiškų atsakymų nepateikia. Hierarchinė analizė gana paranki klasterizuojant mažiau statistiškai apdorotus kintamuosius (ΔV , ΔC , $\%V$, iš dalies ir $\text{varco}\Delta V$, $\text{varco}\Delta C$). Dviejų žingsnių analizę klasifikuojant kalbas pagal ritmą reikėtų taikyti labai atsargiai, nes jos rezultatai prieštaringi.

Lietuvių ir latvių kalbos pagal visus analizuotus kintamuosius patenka į tą patį klasterį kartu su daugeliu kirtinio ritmo kalbų. Todėl galima būtų daryti atsargią išvadą, kad abiejų gyvųjų baltų kalbų ritmas panašus ir jį lemia kirtis. Tai kirtinio ritmo kalbos.

ANNOTATION

This article aims to answer two questions: 1) whether the cluster analysis is handy for finding typological rhythm similarities and on the basis of it so far unanalysed languages can be classified 2) which cluster – stress-timed or syllable-timed languages – Lithuanian and Latvian languages fall into.

For this study five native speakers (Latvian, Lithuanian, Swedish, Icelandic and Chinese) translated and recorded the same text. Text duration was about 1 minute. Records were annotated with the program PRAAT and acoustic correlates of rhythm were calculated using program *Correlatore* developed by Paolo Mairano (2.1 version). Results of other languages estimated by the author of the program Mairano (2011) was used for cluster analysis. The data was entered to the SPSS program and a cluster analysis was performed.

The results show that the hierarchical and two-step cluster analysis does not provide unambiguous results. Hierarchical analysis is pretty handy to cluster less statistically processed variables (ΔV , ΔC , $\%V$, partly $\text{varco}\Delta V$, $\text{varco}\Delta C$). Two-step analysis should be used with caution because of its controversial results.

Lithuanian and Latvian languages from all the variables fall into the same cluster together with a number of stress-timed languages. Therefore, it can be carefully concluded that rhythm of the two Baltic languages is similar and caused by stress. They are stress-timed languages.

1. TRUMPA KALBOS RITMO TYRIMŲ APŽVALGA

Kenneth Lee Pike (1945) vienas iš pirmųjų plačiau aptarė neeiluotos kalbos ritmiškumo tipologinius skirtumus. Jis skyrė: 1) kirtinio ritmo (ang. *stress-timed*) kalbas, kurių ritmo pagrindas yra kirtis (pavyzdžiui, anglų, vokiečių kalbos), 2) skiemėninio ritmo (ang. *syllable-timed*) kalbas, kurių ritmiškumas paremtas

panašios kiekybės skiemenų vartoseną (pavyzdžiui, prancūzų, ispanų kalbos). Prie šių dviejų tipologinių grupių kiek vėliau pridėta ir trečioji – morinio ritmo grupė, jai priklausančių kalbų ritmo pagrindą sudaro mora (pavyzdžiui, japonų, tamilų kalbos).

Daugelis XX a. pabaigos tyrimų paremti Pike siūlyta kalbų ritmo tipologinės klasifikacijos hipoteze. Matuojant ir lyginant akustinę skiemenų ir tarpkirtinių intervalų trukmę, buvo siekiama pagrįsti dviejų ritmo grupių egzistavimą ir jų skirtumus (Abercrombie 1967; Roach 1982). Deja, eksperimentiniai tyrimai garsinės kalbos segmentų izochroniškumo – kartojimosi lygiais intervalais ar apylygės jų kiekybės – nepatvirtino. Tad bandyta ieškoti kitų kalbos ritmiškumo rodiklių. Rebecca McKeever Dauer (1983), remdamasi tuo, kad fonologinė kalbos sistema tiesiogiai lemia ritmą, siūlė skirstyti kalbas į ritmo grupes pagal tam tikrų fonologinių ypatybių buvimą arba nebuvimą. Tai balsių redukcija ir skiemens struktūros sudėtingumas. Remdamiesi šia idėja ir pritaikę statistinius metodus, Franckas Ramusas, Marina Nespor ir Jacques Mehleris (1999) bandė nustatyti akustinius kalbų ritmo koreliatus (kintamuosius, gaunamus apskaičiuojant balsinių ir priebalsinių intervalų trukmę): ΔV – standartinių balsinių intervalų trukmės nuokrypį, ΔC – standartinių priebalsinių intervalų trukmės nuokrypį, $\%V$ – balsinių intervalų trukmės proporciją. Minėti autoriai hipotetiškai spėjo, kad kirtinio ritmo kalboms būdingas didesnis balsių trukmės variantiškumas ir sudėtingesnė skiemens struktūra, todėl ΔV ir ΔC turi būti aukštesni, o $\%V$ – žemesnis nei skiemeninio ritmo kalbų. Deja, jų analizuotų konkrečių kalbų rezultatai tai patvirtino tik iš dalies.

Vėlesniuose tyrimuose atkreiptas dėmesys į galimą įvairių kalbinių ir nekalbinių veiksnių įtaką ritmui. Siekdamas normalizuoti tempo skirtumus, Esther Grabe ir Ling Ee Low (2002) skaičiavo porinius variantiškumo indeksus: nPVI – balsinių intervalų, rPVI – priebalsinių intervalų indeksą. Tikėtasi, kad kirtinio ritmo kalboms būdingas didelis balsių ir priebalsių trukmės variantiškumas, jų nPVI ir rPVI yra aukštas, o skiemeninio ritmo kalbų šie indeksai žemi.

Išvengti tempo įtakos trukmei siekė ir Volkeris Dellwo su Petra Wagner (2003). Jie balsinių ir priebalsinių intervalų trukmės standartinių nuokrypį padaugino iš 100 ir apskaičiavo variacijos koeficientus: $\text{varco}\Delta V$ (balsinių intervalų) bei $\text{varco}\Delta C$ (priebalsinių intervalų). Tačiau nei poriniai variantiškumo indeksai, nei variacijos koeficientai nepadeda vienareikšmiškai nustatyti ritmo tipologinių panašumų ir suskirstyti kalbų į tipologines klases. Vis dėlto aptarta ritmo parametrų nustatymo metodika gana ilgai buvo, o ir dabar dar tebėra populiari, nors ir susilaukė pagrįstos kritikos (pavyzdžiui, Arvaniti 2012)².

² Plačiau apie aptartus kalbos ritmo tyrimo metodus ir galimybę pritaikyti juos aprašant lietuvių kalbos ritmą žr. Kazlauskienė 2013.

2. TYRIMO METODOLOGIJA

Šiam tyrimui pasirinkta išbandyti naują būdą – eksperimentiškai patikrinti akustinių ritmo koreliatų ir apskritai kalbos ritmo panašumą taikant klasterinės analizės metodą. Statistikoje jis naudojamas norint nustatyti objektų ar reiškinių panašumą ir juos suskirstyti į klasterius (t. y. panašių objektų grupes). Klasterinės analizės tikslas – sugrupuoti tiriamuosius dydžius taip, kad skirtumai klasterio viduje būtų kuo mažesni, o tarp klasterių – kuo didesni (Vaitkevičius, Saudargienė 2006).

Tikėtasi, kad, atlikus klasterinę analizę, bus išsiaiškinti du dalykai, kurie ir yra šio tyrimo tikslas.

1. Ar Paolo Mairano (2011) ir šiam tyrimui apskaičiuoti akustiniai švedų, islandų ir kinų kalbų ritmo koreliatai patenka į tuos pačius klasterius? Ar Mairano ir mūsų nustatyti kirtinio ritmo švedų bei islandų kalbų koreliatai skiriasi nuo skiemeninio ritmo kinų kalbos? Jei į šiuos klausimus bus atsakyta teigiamai, vadinasi, klasterinė analizė tikrai paranki ieškant ritmo tipologinių panašumų ir ja remiantis galima klasifikuoti iki šiol neanalizuotas kalbas.

2. Antrasis rūpimas klausimas, į kurį klasterį – skiemeninio ar kirtinio ritmo kalbų – patenka lietuvių ir latvių kalbos?

Akstiniai švedų, islandų ir kinų kalbų ritmo koreliatai jau yra skaičiuoti kitų autorių (šiam tyrimui pasiremta duomenimis iš visiems prieinamo šaltinio internete – Mairano 2011). Pirmosios dvi kalbos priskirtos prie kirtinio ritmo, kinų – prie skiemeninio ritmo grupės. Lietuvių ir latvių kalbų ritmas pagal šią metodiką dar nenagrinėtas. Ankstesniuose lietuvių kalbos ritmiškumo tyrimuose ritmo pagrindu laikytas kirtis, bet skaičiuota ne akustinė vienetų trukmė, o kirčio distribucija įvairiuose segmentuose. Eiliuotos kalbos ritmą šitaip tyrė Juozas Girdzijskas (1966, 1979) ir Leonardas Sauka (1974, 1978), o prozos ir sakininės kalbos ritmą – Valerija Ramonaitė (1980, 2006), Bonifacas Stundžia (1991) ir Asta Kazlauskienė, Inga Fokaitė (2005).

Užuominų apie sakininės latvių kalbos ritmiškumą jau būta. Zinny Sans Bond, Dace Markus ir Verna Stockmal (2003) lygino gimtakalbių latvių ir gimtakalbių rusų, kalbančių latviškai, prozodiją: kontrastinę balsių trukmę, kirtį ir ritmą. Jų straipsnyje nepateikta išsami latvių kalbos ritmo analizė ir interpretacija, nors skaičiuoti Ramuso, Nespore, Mehlerio (1999) ir Grabe bei Low (2002) siūlyti dydžiai. Bond, Markus ir Stockmal (2003) tyrime daryta prielaida, kad latvių kalbai būdingas skiemeninis ritmas.

Šiam tyrimui pasirinktas standartinis Tarptautinės fonetikų asociacijos (žr. http://alt-usage-english.org/north_txt.html) siūlomas tekstas (ang. *The North Wind and the Sun*, lie. *Šiaurės vėjas ir saulė*), kurį naudoja ne tik ritmologai, bet ir visi fonetikai. Analizuojant kitokio pobūdžio medžiagą, būtų nemetodiška šio tyrimo rezultatus lyginti su kitų kalbų duomenimis, gautais remiantis minėtu tekstu.

Penki gimtakalbiai (latvis, lietuvė, švedas, islandė ir kinė)³ tekstą išvertė ir perskaitė. Informantų kalba įrašyta diktofonu ramioje patalpoje. Kiekvienas įrašas yra kiek trumpesnis nei viena minutė.

Tyrimo medžiaga tvarkyta ir empiriniai duomenys išgauti keliais etapais:

1. Konsultuojantis su kiekvienos kalbos informantu, programa PRAAT buvo anotuoti įrašai. Anotuojant remtasi ir oscilogramos vaizdu, ir klausa. Sutvarkytas programos tekstinių anotacijų tinklelis (ang. *TextGride*) išsaugotas **txt* formatu.

2. Naudojant Turino (Italija) Arturo Genre eksperimentinės fonetikos laboratorijoje Mairano (2011) sukurtą programą *Correlatore* (2. 1 versija), buvo apskaičiuoti mūsų atliktų įrašų akustinių ritmo koreliatų statistiniai dydžiai (rezultatai pateikiami 1 lent.). Klasterinei analizei, kaip jau minėta, panaudoti duomenys iš Mairano (2011). Jie reikalingi tam, kad būtų galima tinkamai atlikti klasterinę analizę, kurios rezultatai gerokai patikimesni, kai grupuojamas didesnis kiekis duomenų.

1 LENTELE. Akustinių ritmo koreliatų statistinių dydžių reikšmės⁴

Kalba	ΔC	%V	ΔV	varco ΔV	varco ΔC	nPVI	rPVI
Lietuvių	52	54	54	54	66	57	57
Latvių	55	52	58	55	65	70	65
Švedų	55	44	36	39	48	45	68
Švedų 2	65	49	51	50	66	60	82
Islandų	60	40	33	45	56	45	63
Islandų 2	52	47	41	50	60	51	56
Kinų	44	51	44	44	47	48	54
Kinų 2	48	62	74	51	62	51	53
Japonų	31	51	39	47	40	43	32
Portugalų	54	49	55	48	44	57	56
Prancūzų	40	49	40	45	41	45	44
Suomių	47	48	45	47	46	45	55
Graikų	37	47	34	50	48	52	42
Vokiečių	66	46	54	56	60	58	65
Estų	54	46	39	43	51	43	53
Čekų	58	45	51	51	49	45	61
Turkų	58	45	38	43	55	46	67

³ Visi informantai – aukštąjį filologinį išsilavinimą turintys žmonės, dirbantys arba tęsiantys studijas universitetuose: latvis ir islandė – dėstytojai, švedas – personalo vadovas, lietuvė ir kinė – magistrantės.

⁴ Čia ir toliau mūsų tyrimo įrašai žymimi su indeksu 2, pvz.: švedų 2, o Mairano duomenys – be indekso.

Kalba	ΔC	%V	ΔV	varco ΔV	varco ΔC	nPVI	rPVI
Arabų	55	45	47	56	55	47	64
Italų	54	44	44	53	49	48	62
Olandų	46	43	45	62	50	68	53
Danų	61	42	40	46	52	50	66
Ispanų	52	42	30	42	53	42	60
Rusų	55	41	34	43	48	45	63
Anglų	59	40	48	62	51	61	70
Lenkų	66	36	32	41	49	41	72

3. Duomenys suvesti į SPSS programą (apie ją plačiau žr. <http://www.insol.lt/statistics/>, Vaitkevičius, Saudargienė 2006) ir atlikta jų klasterinė analizė.

Nagrinėjant lietuvių ir latvių kalbas SPSS programa, buvo atlikta hierarchinė ir dviejų žingsnių klasterinė analizė. Taikant hierarchinę analizę, iš anksto nėra žinomas klasterių skaičius. Klasteriai sudaromi hierarchiniu principu – visi duomenys laikomi vienu klasteriu, sudarytu iš mažesnių grupių. Jie gali būti nustatomi vieną bendrą klasterį skaidant į mažesnius arba mažus klasterius jungiant į didesnius.

Dviejų žingsnių analizė nuo aptartosios skiriasi tuo, kad ją taikant klasterių skaičius yra žinomas iš anksto (Vaitkevičius, Saudargienė 2006). Darbe pasirinkti du klasteriai, nes nagrinėjamos dvi kalbų ritmo grupės – kirtinio ir skiemeninio. Trečioji, morinio ritmo, grupė neįtraukta į analizę, nes nekelta hipotezė, kad baltų kalbos galėtų priklausyti šiai grupei. Atliekant klasterinę analizę, objektai periskirstomi tol, kol panašumai klasterio viduje tampa gerokai didesni už panašumus tarp klasterių (Vaitkevičius, Saudargienė 2006). Švedų, islandų ir kinų kalbų rezultatai tikrinami taikant tik dviejų žingsnių klasterinę analizę, nes pasirinktas mažas nagrinėjamų objektų skaičius (šešios kalbos). Be to, iš anksto žinomas klasterių skaičius ir kuriai ritmo grupei šios kalbos priskiriamos (dvi ritmo grupės ir dviejų tyrimų rezultatai).

3. ŠVEDŲ, ISLANDŲ IR KINŲ KALBŲ AKUSTINIŲ KORELIATŲ DVIEJŲ ŽINGSNIŲ KLASTERINĖ ANALIZĖ

Standartinis priebalsių trukmės nuokrypis (ΔC)
ir balsinių intervalų trukmės procentas (%V)

Siekiant išsiaiškinti, ar Mairano (2011) ir šiame tyrime įrašytų švedų, islandų bei kinų kalbų %V ir ΔC dydžiai yra panašūs ir patenka į tą pačią grupę, buvo atlikta

dviejų žingsnių klasterinė analizė. Pirmiausia apskaičiuotos kiekvieno klasterio kiekybinių kintamųjų vidurkių reikšmės ir jų standartinis nuokrypis (žr. 2 lent.).

2 LENTELĖ. Klasterių centrai pagal balsinių intervalų trukmės procentą ir standartinį priebalsių trukmės nuokrypį

		%V		ΔC	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	45,00	3,916	58,00	5,715
	2	56,50	7,778	46,00	2,828
	Bendras	48,83	7,521	54,00	7,720

Iš lentelės matyti, kad pirmojo klasterio ΔC vidurkis aukštesnis nei antrojo, o kintamojo %V, atvirkščiai, – antrojo klasterio aukštesnis, bet šiuo atveju yra gana didelis ir standartinis nuokrypis. Vadinasi, pirmajam klasteriui būdinga sudėtingesnė skiemens struktūra (t. y. daugiau priebalsių skiemenyje (ΔC aukštesnis)), kuri lemia ir mažesni balsinių intervalų procentą (%V).

Šios analizės rezultatai rodo, kad panašiausius %V bei ΔC turi ir vieną klasterį sudaro kirtinio ritmo grupei priklausančios islandų (%V = 40, ΔC = 60), islandų 2 (%V = 47, ΔC = 52), švedų (%V = 44, ΔC = 55), švedų 2 (%V = 49, ΔC = 65) kalbos, o skiemeninio ritmo kinų (%V = 51, ΔC = 44) ir kinų 2 (%V = 62, ΔC = 48) pavyzdžiai suformavo kitą klasterį. Vadinasi, %V ir ΔC yra gana patikimas rodiklis, pagal kurį analizuotos kalbos iš tiesų skiriasi ir kuris nepriklauso nuo tyrėjų ar informantų objektyvių (kalbėjimo tempo, intonacijos ir kt.) ar subjektyvių (įrašo kokybės, anotavimo skirtumų ir kt.) veiksnių.

Variacijos koeficientai

Gana nemažas skirtumas tarp abiejų klasterių centrų yra ir pagal variacijos koeficientus. 3 lentelėje matyti, kad antrojo klasterio varcoΔV ir varcoΔC vidurkiai didesni.

3 LENTELĖ. Klasterių centrai pagal variacijos koeficientus

		varcoΔV		varcoΔC	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	42,67	3,215	50,33	4,933
	2	50,33	0,517	62,67	3,055
	Bendras	46,50	4,680	56,50	7,688

Deja, ši kartą rezultatai labai netikėti: panašiausius varco Δ V ir varco Δ C dydžius turi ir pirmąjį klasterį sudaro Mairano (2011) analizuotos islandų (varco Δ V = 45, varco Δ C = 56), švedų (varco Δ V = 39, varco Δ C = 48) ir kinų (varco Δ V = 44, varco Δ C = 47) kalbos. Visi mūsų tiriamieji įrašai patenka į antrą klasterį: islandų 2 (varco Δ V = 50, varco Δ C = 60), švedų 2 (varco Δ V = 50, varco Δ C = 66) ir kinų 2 (varco Δ V = 51, varco Δ C = 62). Taigi variacijos koeficientas, turėjęs būti objektyvesnis tipologiškai klasifikuojant kalbas, kelia daugiau painiavos ir leidžia spėti, kad jo rezultatai labai priklauso nuo konkrečios kalbos anotavimo patirties. Mūsų įrašų variacijos koeficientai panašūs, o tai leidžia manyti, kad anotuojant galbūt nepakanka tik konsultuotis su gimtakalbiais informantais, bet patį anotavimą reikia atlikti gimtakalbiams fonetikams.

Poriniai variantiškumo indeksai

Skirtumas tarp abiejų klasterių centrų pagal nPVI ir rPVI gana nemažas (žr. 4 lent.).

4 LENTELĖ. Klasterių centrai pagal porinius variantiškumo indeksus

		nPVI		rPVI	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	48,00	3,000	58,80	6,458
	2	60,00	.	82,00	.
	Bendras	50,00	5,586	62,67	11,094

Tačiau kalbų susigrupavimas dar dviprasmiškesnis nei nagrinėjant variacijos koeficientus. Pirmą klasterį sudaro visos kalbos, išskyrus šiam tyrimui analizuotą švedų 2 kalbą (jos rPVI labai aukštas – 82, nors nPVI panašus į kitų kalbų – 60) – ji vienintelė pateko į antrąjį klasterį. Kitų kalbų poriniai variacijos indeksai labai panašūs: islandų (nPVI = 45, rPVI = 63), islandų 2 (nPVI = 51, rPVI = 56), kinų (nPVI = 48, rPVI = 54), kinų 2 (nPVI = 51, rPVI = 53), švedų (nPVI = 45, rPVI = 68). Šiuo atveju akivaizdu, kad dviejų žingsnių klasterinė analizė nėra naudinga, nes pirmąjį klasterį reikia skaidyti dar į smulkesnes grupes.

Apibendrinant trijų kalbų dviejų žingsnių klasterinės analizės rezultatus, galima sakyti, kad standartinis priebalsių trukmės nuokrypis (Δ C) ir balsinių intervalų trukmės procentas (%V) yra gana patikimas rodiklis grupuojant kalbas į klases. Variacijos koeficientų ir porinių variantiškumo indeksų dviejų žingsnių klasterinė analizė šiuo atveju kelia daugiau klausimų, nei pateikia atsakymų.

4. LIETUVIŲ IR LATVIŲ KALBŲ RITMAS
REMIANTIS KLASTERINE ANALIZEStandartiniai trukmės nuokrypiai (ΔC , ΔV)
ir balsinių intervalų trukmės procentas (%V)

Norint išsiaiškinti, kuriai tipologinei ritmo grupei priklauso latvių ir lietuvių kalbos, pirmiausia buvo atlikta hierarchinė klasterinė analizė, kuria siekta patikrinti, kaip pagal ΔV , %V ir ΔC grupuojasi kalbos ir į kurį klasterį patenka lietuvių ir latvių kalbos. Šiai analizei naudoti ir kitų kalbų duomenys, pateikti Mairano darbe (2011: 170, 179, 199).

Gauti hierarchinės klasterinės analizės rezultatai rodo, kad lietuvių, latvių, vokiečių, arabų, anglų, čekų, italų ir portugalų kalbos turi panašiausius ΔV , %V ir ΔC dydžius, todėl sudaro vieną klasterį, kurio kalbų ΔV yra nuo 44 iki 58, %V – nuo 40 iki 54, ΔC – nuo 52 iki 66. Šiame klasteryje tik vienintelė italų kalba yra laikoma artimesne skiemeniniam ritmui (kaip matyti, klasterinė analizė lyg ir leidžia tuo abejoti), visos kitos kalbos (išskyrus, žinoma, analizuojamas baltų kalbas, kurių ritmo rūšis nežinoma) priklauso kirtinio ritmo kalboms.

Antrąją grupę suformavo daugelis skiemeninio ritmo kalbų, bet į ją pateko nemažai tų, kurias linkstama laikyti kirtinio ar mišraus ritmo kalbomis (žr. 5 lent.). Šių kalbų ΔV yra nuo 30 iki 45, %V – nuo 36 iki 51, ΔC – nuo 37 iki 66. Žvelgiant į rezultatus, viena yra neabejotina: į pirmąjį klasterį nepateko nei viena prototipinė skiemeninio ritmo kalba (prancūzų, ispanų, graikų), o į antrąjį pateko tik dvi prototipinės kirtinio ritmo kalbos (rusų, olandų). Kaip matyti iš 5 lentelės, lietuvių ir latvių kalbos patenka į pirmąjį klasterį ir yra panašesnės į prototipines kirtinio ritmo kalbas (kalbų akustinių koreliatų duomenys pateikti 1 lent.).

5 LENTELE. Hierarchinės klasterinės analizės rezultatai pagal standartinius nuokrypius ir balsių proporciją⁵

Klasterio numeris	Ritmo grupės ir kalbos		
	Kirtinis ritmas	Skiemeninis ritmas	Nenustatyta
1	anglų, vokiečių, arabų, portugalų, čekų	italų	lietuvių, latvių
2	rusų, olandų, islandų, turkų, danų, lenkų, švedų	ispanų, prancūzų, graikų, kinų, estų, suomių	

⁵ Šioje ir kitose lentelėse paryškintos prototipinės konkrečios ritmo grupės kalbos, dėl kitų kalbų ritmo diskutuojama (kai kurios iš jų netgi laikomos arba neklasifikuotinomis, arba mišraus ritmo).

Hierarchinės trijų kintamųjų analizės rezultatai leidžia daryti išvadą, kad prototipinės kirtinio ir skiemeninio ritmo kalbos patenka į skirtingus klasterius. Taigi pagal analizuojamus požymius šių grupių kalbos skiriasi, bet ar šie požymiai yra skirtingo kalbos ritmo rodikliai – jau kito tyrimo tikslas.

Atlikus dviejų žingsnių klasterinę analizę (pasirinktas klasterių skaičius, savaime suprantama, yra du), matyti, kad pirmojo klasterio kalbų ΔV vidurkis aukštesnis, o antrojo – žemesnis (žr. 6 lent.). Pirmojo klasterio kalboms būdinga didesnė balsinių intervalų trukmės įvairovė. Kintamojo ΔC vidurkis yra žemesnis pirmajame klasteryje ir kiek aukštesnis antrajame. Tačiau skirtumas nėra labai didelis, ypač atsižvelgus į tai, kad pirmajame klasteryje duomenys yra labai išsibarstę (aukštas standartinis nuokrypis – net 8,15).

6 LENTELĖ. Klasterių centrai pagal standartinius nuokrypius ir balsių proporciją

		%V		ΔV		ΔC	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	47,75	3,388	47,58	6,999	50,67	8,150
	2	41,78	3,032	36,67	5,408	57,78	4,295
	Bendras	45,19	4,377	42,90	8,324	53,71	7,544

Dviejų žingsnių klasterinės analizės rezultatai, pateikti 7 lentelėje, rodo, kad lietuvių ir latvių kalbų ΔV , %V ir ΔC dydžiai yra panašiausi į kirtinių vokiečių, arabų, olandų, portugalų, čekų ir skiemeninių italų, graikų, kinų, suomių bei prancūzų kalbų. Pirmąjį klasterį, kuriam, šio tyrimo duomenimis, priklauso ir baltų kalbos, sudaro 12 (57 %), antrąjį – 9 (43 %) kalbos.

7 LENTELĖ. Dviejų žingsnių klasterinės analizės rezultatai pagal standartinius nuokrypius ir balsių proporciją

Klasterio numeris	Ritmo grupės ir kalbos		
	Kirtinis ritmas	Skiemeninis ritmas	Nenustatyta
1	arabų, vokiečių, olandų, portugalų, čekų	prancūzų, graikų, italų, kinų, suomių	lietuvių, latvių
2	anglų, rusų, islandų, turkų, danų, lenkų, švedų	ispanų, estų	

Matyti, kad šios analizės rezultatai nuo hierarchinės skiriasi tuo, kad į tą patį klasterį kartu su lietuvių ir latvių kalbomis pateko daugiau skiemeninio ritmo

kalbų: italų, kinų, suomių ir net prototipinės prancūzų bei graikų kalbos. Kita prototipinė skiemeninio ritmo kalba (ispanų) pateko į antrąją klasterį. Ši kartą išskaidytos ir prototipinės kirtinio ritmo kalbos: anglų – antrajame, vokiečių – pirmajame klasteryje. Tad analizuojant tris kintamuosius (pačius svarbiausius kalbų ritmo tipologinės klasės rodiklius), tinkamesnė yra hierarchinė, o ne dviejų žingsnių klasterinė analizė.

Variantiškumo koeficientai

Hierarchinė klasterinė analizė pagal kintamuosius $\text{varco}\Delta V$, $\text{varco}\Delta C$ rodo, kad į vieną klasterį patenka lietuvių, latvių ir keturios prototipinės kirtinio ritmo kalbos: arabų, olandų anglų ir vokiečių (žr. 8 lent.). Šių kalbų $\text{varco}\Delta V$ yra nuo 54 iki 62, $\text{varco}\Delta C$ – nuo 50 iki 66. Antrojo klasterio kalbų $\text{varco}\Delta V$ yra nuo 39 iki 53, $\text{varco}\Delta C$ – nuo 41 iki 56.

8 LENTELE. Hierarchinės ir dviejų žingsnių klasterinės analizės rezultatai pagal variantiškumo koeficientus

Klasterio numeris	Ritmo grupės ir kalbos		
	Kirtinis ritmas	Skiemeninis ritmas	Nenustatyta
1	anglų, olandų, vokiečių, arabų	–	lietuvių, latvių
2	rusų, portugalų , islandų, turkų, danų, čekų, lenkų, švedų	ispanų, prancūzų, graikų , estų, italų, kinų, suomių	

Kaip matyti iš dviejų žingsnių klasterinės analizės rezultatų, pirmojo klasterio abiejų variantiškumo koeficientų vidurkiaiai yra aukštesni nei antrojo (žr. 9 lent.).

9 LENTELE. Klasterių centrai pagal variantiškumo koeficientus

		$\text{varco}\Delta V$		$\text{varco}\Delta C$	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	57,50	3,564	57,83	6,911
	2	45,33	3,885	49,07	3,955
	Bendras	48,81	6,743	51,57	6,274

Hierarchinės ir dviejų žingsnių klasterinės analizės rezultatai visiškai sutampa. Klasterius sudaro tos pačios kalbos: viename yra lietuvių, latvių, vokiečių, arabų, anglų ir olandų kalbos, kitame – likusios (žr. 8 lent.).

Hierarchinė ir dviejų žingsnių klasterinė variantiškumo koeficientų analizė aki-vaizdžiai rodo lietuvių, latvių ir kirtinių kalbų ritmo panašumą bei gana tiksliai atskiria prototipines kirtines kalbas nuo skiemeninių.

Poriniai variantiškumo indeksai

Hierarchinė klasterinė analizė pagal porinius variantiškumo indeksus pateikia labai diskutuotinus rezultatus (žr. 10 lent.). Visos kalbos, išskyrus prancūzų ir graikų, patenka į vieną klasterį. Pirmojo klasterio kalbų rPVI yra nuo 53 iki 72, nPVI – nuo 41 iki 70. Antrojo klasterio kalbų rPVI yra nuo 42 iki 44, nPVI – nuo 45 iki 52.

10 LENTELĖ. Hierarchinės klasterinės analizės rezultatai pagal porinius variantiškumo indeksus

Klasterio numeris	Ritmo grupės ir kalbos		
	Kirtinis ritmas	Skiemeninis ritmas	Nenustatyta
1	anglų, olandų, vokiečių, arabų, rusų, portugalų, islandų, turkų, danų, čekų, lenkų, švedų	ispanų, estų, italų, kinų, suomių	lietuvių, latvių
2	–	prancūzų, graikų	

11 lentelėje pateikti dviejų žingsnių klasterinės analizės rezultatai. Atkreiptinas dėmesys į tai, kad abiejų klasterių rPVI vidurkiai yra beveik vienodi, o tai labai leidžia abejoti šio rodiklio patikimumu (ypač turint galvoje aukštą standartinį nuokrypį).

11 LENTELĖ. Klasterių centrai pagal porinius variantiškumo indeksus

		nPVI		rPVI	
		Vidurkis	Standartinis nuokrypis	Vidurkis	Standartinis nuokrypis
Klasteris	1	61,83	5,776	61,00	6,603
	2	45,80	2,883	59,60	8,534
	Bendras	50,38	8,322	60,00	7,893

Pirmąjį klasterį sudaro šešios kalbos (29 %), antrąjį – 15 (71 %) (žr. 12 lent.). Jeigu lygintume su hierarchinės klasterinės analizės rezultatais, matytume, kad pirmajame klasteryje kartu su lietuvių, latvių kalbomis liko vokiečių, anglų, olandų ir portugalų kalbos. Visos skiemeninio ritmo kalbos pateko į antrąjį klasterį.

12 LENTELE. Dviejų žingsnių klasterinės analizės rezultatai pagal porinius variantiškumo indeksus

Klasterio numeris	Ritmo grupės ir kalbos		
	Kirtinis ritmas	Skiemeninis ritmas	Nenustatyta
1	anglų, olandų, vokiečių, portugalų	–	lietuvių, latvių
2	rusų, arabų , islandų, turkų, danų, čekų, lenkų, švedų	ispanų, prancūzų, graikų , estų, italų, kinų, suomių	

Dviejų žingsnių klasterinės analizės rezultatai, kai kintamieji nPVI ir rPVI, rodo lietuvių, latvių ir kirtinio ritmo kalbų ritmo panašumą ir, kaip ir variantiškumo koeficientų analizė, gerai atskiria prototipines kirtines kalbas nuo skiemeninių.

5. IŠVADOS

Apibendrinant bandymą klasterinės analizės metodus pritaikyti tipologinėms ritmo klasėms identifikuoti, galima daryti tokias išvadas.

Klasterinė analizė (ir hierarchinė, ir dviejų žingsnių) vienareikšmiškų rezultatų nepateikia. Hierarchinė analizė gana paranki klasterizuojant mažiau statistiškai apdorotus kintamuosius (ΔV , ΔC , %V, iš dalies ir $\text{varco}\Delta V$, $\text{varco}\Delta C$), dviejų žingsnių analize galbūt galima remtis ir turint sudėtingesnius indeksus (PVI). Vis dėlto dviejų žingsnių analizė šiuo atveju kelia daugiau abejonių jau vien dėl to, kad pasirinktos tik dvi galimybės (arba kirtinis, arba skiemeninis ritmas, negali būti mišrus).

Lietuvių ir latvių kalbos pagal visus kintamuosius patenka į tą patį klasterį kartu su daugeliu kirtinio ritmo kalbų. Todėl būtų galima daryti atsargią išvadą, kad abiejų baltų kalbų ritmas panašus ir jį lemia kirtis. Tai kirtinio ritmo kalbos.

Klasterinė analizė (ypač švedų, islandų, kinų kalbų dviejų skirtingų tyrimų duomenų gretinimas) parodė, kad ritmui tirti ypač svarbus anotavimo preciziškumas. Kita vertus, tai tik bandomasis tyrimas, todėl šios išvados aktualios ne dėl pačios tipologinės kalbų ritmo klasifikacijos ar kalbų ritmiškumo rodiklių. Jos gali būti tolesnių tyrimų gairės arba dar vienas pretekstas suabejoti griežtai apibrėžtų kalbos ritmo klasių egzistavimu.

LITERATŪRA

Abercrombie David 1967: *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Arvaniti Amalia 2012: The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40, 351–373.

- Bond Zinny Sans, Markus Dace, Stockmal Verna 2003: Prosodic and Rhythmic Patterns Produced By Native and Non-native Speakers of a Quantity-Sensitive Language. *15th International Congress of Phonetic Sciences*. Barcelona, 527–530.
- Dauer Rebecca McKeever 1983: Stress-timing and syllable-timing reanalysed. *Journal of Phonetics* 11, 51–62.
- Dellwo Volker, Wagner Petra 2003: Relationships between language rhythm and speech rate. *Proc. of the International Congress of Phonetic Science*, 471–474.
- Girdzijauskas Juozas 1966: *Lietuvių eilėdara. Silabinės-toninės sistemos susiformavimas*. Vilnius: Vaga.
- Girdzijauskas Juozas 1979: *Lietuvių eilėdara: XX a.* Vilnius: Mokslas.
- Grabe Esther, Low Ee Ling 2002: Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in Laboratory Phonology* 7, 1–16.
- Kazlauskienė Asta, Fokaitė Inga 2005: Dėl lietuvių kalbos teksto prozodijos. *Žmogus ir žodis* 7(1), 35–38.
- Kazlauskienė Asta 2013: Kai kurie akustiniai lietuvių kalbos ritmiškumo parametrai. *Baltistica* 48(1), 19–36.
- Mairano Paolo 2011: *Rhythm Typology: Acoustic and Perceptive Studies*. Doctoral dissertation. Turin: University of Turin.
- Pike Kenneth Lee 1945: *The Intonation of American English*. Ann-Arbor: University of Michigan Press.
- Pukevičiūtė Agnė 2013: *Akustiniai lietuvių ir latvių kalbų ritmo koreliatai*. Magistro darbas. Kaunas: Vytauto Didžiojo universitetas.
- Ramonaitė Valerija 1980: Prozos kalbos ritmas. Konstruktyvieji faktoriai ir pagalbiniai formantai. *Literatūra ir kalba* 16, 38–65.
- Ramonaitė Valerija 2006: Pirminis Juozo Baltušio „Parduotų vasarų“ ritmas. *Acta humanitarica universitatis Saulensis*, 307–333.
- Ramus Franck, Nespour Marina, Mehler Jacques 1999: Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3), 265–292.
- Roach Peter 1982: On the Distinction between ‘Stress-timed’ and ‘Syllable-timed’ Languages. – David Crystal. *Linguistic controversies*. London: Edward Arnold, 73–79.
- Sauka Leonardas 1974: Lietuvių folklore eilėtyros ir teorijos klausimai. – *Literatūra ir kalba* 13, 192–207.
- Sauka Leonardas 1978: *Lietuvių liaudies dainų eilėdara*. Vilnius: Vaga.
- Stundžia Bonifacas 1991: Kirtis tekste. *Žodžiai ir prasmės* 1, 86–92.
- Vaitkevičius Raimundas, Saudargienė Aušra 2006: *Statistika su SPSS psichologiniuose tyrimuose: mokomoji knyga*. Kaunas: VDU leidykla.
- Vaitkevičius Raimundas, Saudargienė Aušra 2012: *Daugiamąčiai duomenų analizės metodai psichologijoje*. Kaunas.

PRIEDAS

1. Tiriamieji tekstai lietuvių kalba

Vieną dieną Šiauriniam vėjui ir Saulei besiginčijant, kuris iš jų stipresnis, pasirodė keliautojas, vilkintis apsiaustą. Tada jie susitarė, kad tas, kuris privers keliautoją nusivilkti apsiaustą, bus laikomas stipresniu.

Pirmiausia Šiaurinis vėjas pūtė taip stipriai, kiek tik galėjo, bet kuo stipriau jis pūtė, tuo tvirčiau keliautojas apsigaubdavo apsiaustą; galiausiai Šiaurinis Vėjas pasidavė.

Tada Saulė pradėjo kaitinti ir keliautojas tuoj pat nusivilkio apsiaustą. Taip Šiaurinis vėjas turėjo pripažinti, kad Saulė stipresnė už jį.

2. Tiriamieji tekstai latvių kalba

Vienu dienu Ziemeļu vējam un Saulei strīdoties par to, kurš no viņiem ir stiprāks, nāca ceļotājs, ietīnīes mēteli. Viņi savā starpā vienojās, ka tas, kurš pirmais liks ceļotājam novilkst savu mēteli, tiks uzskatīts par stiprāko.

Ziemeļvējš pirmais pūta tik stipri, cik vien varēja, bet, jo stiprāk viņš pūta, jo stingrāk ceļotājs ietīnās savā mētelī; galu galā Ziemeļvējš tomēr padevās.

Tad Saule sāka stipri sildīt un ceļotājs uzreiz novilkā savu mēteli. Tā Ziemeļvējam bija jāatzīst, ka Saule bija stiprāka par viņu.

3. Tiriamieji tekstai švedų kalba

Nordanvinden och Solen tvistade en gång om vem av dem som var starkast. Just då kom en vandrare vägen fram insvept i en varm kappa. De kom då överens om att den som först kunde få vandraren att ta av sig kappan, han skulle anses vara starkare än den andra.

Då blåste nordanvinden så hårt han nånsin kunde, men ju hårdare han blåste desto tätare svepte vandraren kappan om sig, och till slut gav Nordanvinden upp försöket.

Då lät Solen sina strålar skina helt varmt och genast tog vandraren av sig kappan och så var Nordanvinden tvungen att erkänna att solen var den starkaste av de två.

4. Tiriamieji tekstai islandų kalba

Dag einn þegar Norðanvindurinn og Sólin voru að rífast um það hvor þeirra væri sterkari, birtist ferðamaður vafinn inn í frakka. Þeir voru sammála um að sá sem gæti blásið frakkann af ferðamanninum myndi teljast sterkari en hinn.

Norðanvindurinn blés eins fast og hann gat, en því fastara sem hann blés, því þéttara vafði ferðamaðurinn kápunni um sig, og að lokum gafst Norðanvindurinn upp á að reyna.

Þá fór Sólin að skína, og undireins fór ferðamaðurinn úr frakkanum. Þá varð Norðanvindurinn að viðurkenna að Sólin væri sterkari en hann var.

6. Tiriamieji tekstai kinų kalba

You yi tian, dang yi ge guo zhe da yi de lv xing zhe chu xian de shi hou, bei feng he tai yang zheng zai zheng chao jiu jing shui geng jiang. Ta men tong yi shui xian rang lv xing zhe ba wai tao tuo diao jiu ren wei shui bi ling yi ge qiang.

Sui hou, bei feng jiu jing ta zui da de nu li lai chui feng, dan shi, ta chui de yue da, lv xing zhe jiu ba wai tao guo de yue jin. Zui hou, bei feng fang qi le.

Jie xia lai, tai yang kai shi fa chu re liang, lv xing zhe li ji jiu tuo xia le wai tao. Suo yi, bei feng bu de bu cheng ren tai yang bi ta geng qiang.

Cluster Analysis of Typological Groups of Language Rhythm

SUMMARY

This article aims to answer two questions: 1) whether the cluster analysis handy finding rhythm typological similarities, and it can be classified on the basis of it so far unanalysed languages 2) in which a cluster – stress-timed or syllable-timed languages – Lithuanian and Latvian languages fall into.

Standard text, proposed by the International Phonetics Association (see http://alt-usage-english.org/north_txt.html) (Eng. *The North Wind and the Sun*, Lith. *Šiaurės vėjas ir saulė*) was selected for this study. Five native speakers (Latvian, Lithuanian, Swedish, Icelandic and Chinese) translated and recorded it (each record was much shorter than 1 min.). The records were annotated by program PRAAT and acoustic correlates of rhythm of records that we have done have been calculated using program *Correlator* (2. 1 version) which was developed by Pado Mairano (2011) in Arturo Genre experimental phonetics laboratory of Turin (Italy). It was found that Lithuanian $\Delta C=52$, $\%V=54$, $\Delta V=54$, $\text{varco}\Delta V=54$, $\text{varco}\Delta C=66$, $nPVI=57$, $rPVI=57$; Latvian $\Delta C=55$, $\%V=52$, $\Delta V=58$, $\text{varco}\Delta V=55$, $\text{varco}\Delta C=65$, $nPVI=70$, $rPVI=65$; Swedish $\Delta C=65$, $\%V=49$, $\Delta V=51$, $\text{varco}\Delta V=50$, $\text{varco}\Delta C=66$, $nPVI=60$, $rPVI=82$; Icelandic $\Delta C=52$, $\%V=47$, $\Delta V=41$, $\text{varco}\Delta V=50$, $\text{varco}\Delta C=60$, $nPVI=51$, $rPVI=56$; Chinese $\Delta C=48$, $\%V=62$, $\Delta V=74$, $\text{varco}\Delta V=51$, $\text{varco}\Delta C=62$, $nPVI=51$, $rPVI=53$.

Data of other languages estimated by the author of the program Mairano (2011) was used for cluster analysis. The data was entered to the SPSS program and a cluster analysis was performed.

The results show that the cluster analysis (both hierarchical and two-steps) does not provide unambiguous results. Hierarchical analysis is pretty handy to cluster less statistically processed variables (ΔV , ΔC , $\%V$, partly $\text{varco}\Delta V$, $\text{varco}\Delta C$), a two-step analysis might be possible to rely on, and with more complex indices (PVI). However, a two-step analysis is begging the question, because it does not leave the third option (stress-timed or syllable-timed, cannot be mixed).

Lithuanian and Latvian languages according to all the variables fall into the same cluster together with a number of stress-timed languages. Therefore, it can be carefully concluded that rhythm of the two Baltic languages is similar and caused by stress. They are stress-timed languages.

Cluster analysis (in particular, Swedish, Icelandic, Chinese Language comparison between the two types of data) has shown that the annotation precision is particularly important in the researches of rhythm. On the other hand, the findings are relevant not for the typological classification of languages' rhythm and not for the indicators of languages' rhythm. They can be as guidelines of further research or another excuse to question existence of strictly defined classes of speech rhythm.

Įteikta 2013 m. liepos 11 d.

AGNĖ PUKEVIČIŪTĖ

Vytauto Didžiojo universitetas

Kristijono Donelaičio g. 58, LT-44248 Kaunas, Lietuva

agne.pukeviciute@fc.vdu.lt

ASTA KAZLAUSKIENĖ

Vytauto Didžiojo universitetas

Kristijono Donelaičio g. 58, LT-44248 Kaunas, Lietuva

a.kazlauskiene@hmf.vdu.lt