et

Kahekümne esimene rahvusvaheline lingvistikaolümpiaad

Brasília (Brasiilia), 23.–31. juuli 2024

Meeskonnavõistluse ülesanne

Leksikostatistika on grupp meetodeid, mida kasutatakse selleks, et hinnata, kui lähedased on keeled üksteisele sõnavara poolest. Neid meetodeid kohaldatakse tavaliselt pikkadele sõnaloeteludele, mida spetsialistid käsitsi kommenteerivad, et näidata, kas antud sõnapaar on eeldatavalt samast allikast pärit. Mõnikord kasutavad keeleteadlased siiski leksikostatistilisi meetodeid automaatselt märgendatud sõnaloeteludele. Üks selline meetod põhineb *konsonandiklasside* mõistel, mille nõukogude-iisraeli keeleteadlane Aharon Dolgopolski võttis kasutusele aastal 1964.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ʁ	Y.	j ç (tüve alguses)	M.	m ɱ
T.	t d ɗ θ ð ʈ ɖ	R.	r ɾ ɽ ɹ l ɬ ɮ ɭ ʎ ɬ	W.	w ʍ (tüve alguses)	N.	n ɳ ɲ ŋ
S.	s z ʃ ʒ ʂ ʐ ɕ ʑ c ɟ					Q.	t͡ɬ d͡ɮ
H.	ħ ʕ ʜ ʢ ʔ h ɦ ʔ, vokaalid ja j ç w ʍ (välja arvatud tüve alguses)						

Dolgopolski konsonandiklassid

All leiate märkustega varustatud sõnaloetelude väljavõtted maailma erinevatest keelkondadest. Märkused on esitatud alaindeksitega. Nende loetelude põhjal on konstrueeritud keelepuud, kasutades nn *StarlingNJ* algoritmi kahte lihtsustatud meetodit, ning igale sõnale on määratud *stabiilsusindeks*. Ülemises osas esitatud puud ja stabiilsusindeksid põhinevad käsitsi märgendatud sõnaloeteludel ja alumises osas esitatud puud ja stabiilsusindeksid põhinevad automaatselt märgendatud loeteludel. Iga sõnaloetelu jaoks on olemas kaks konstrueeritud puud, vastavalt algoritmi mõlemale versioonile: Algoritm A ja Algoritm B. Pange tähele, et mõnikord on mitu võimalikku puud, mis vastavad ühele sõnaloetelule; sellisel juhul on üks puu valitud juhuslikult. Igale sõlmpunktile on määratud leksikostatistiline kaugus. Mida suurem on kaugus, seda lähedasemalt on keeled omavahel suguluses. Täpsem mõiste oleks siis „vastupidine leksikostatistiline kaugus“, mitte „leksikostatistiline kaugus“. Et asja lihtsamaks teha, kasutame siin ülesandes mõistet „leksikostatistiline kaugus“.

Nii stabiilsusindeksid kui ka leksikostatistilised kaugused on ümardatud kahe kümnendkohani. Kui kolmas number pärast koma on väiksem kui 5, ümardatakse allapoole; vastasel juhul ümardatakse ülespoole. Näiteks on 2,836 ümardatud 2,84 peale, 0,705 0,71 peale ja 0,703 0,70 peale. Ümardamine kehtib ainult nende väärtuste kohta, mida näidatakse inimestest lugejatele. Seega arvuti „näeb“ algoritmide rakendamise käigus ümardamata väärtusi.

Pange tähele, et mõned sõnad on teadaolevalt või arvatavasti laenatud teistest keeltest. Näiteks kadiveu keele **jokːi** ‘sool’ sõna on laenatud guaranikeelsest **jukɨ** ning ’iipay (Mesa Grande murre) keele **ʔaːnʲ** ‘aasta’ sõna on laenatud hispaaniakeelsest **ˈaɲo**.

Mõnel juhul on sõnade loeteludes ühe tähenduse jaoks esitatud mitu sünonüümi, mis on üksteisest eraldatud komaga. Üks näide on ‘jalg’ vehhose keeles.

Allpool esitatud andmetes on kõik eesliited eraldatud märgiga „=“ ja kõik järelliited märgiga „-“. Mõnda sõna kasutatakse ainult eesliitega koos. Need algavad märgiga „=“.

Andmed on transkribeeritud rahvusvahelise foneetilise tähestiku (IPA) abil. ˈ = pearõhk, ˌ = kaasrõhk (nõrgem kui pearõhk), ◌ː = pikk häälik, ◌̆ = ülilühike häälik, X͡Y = X ja Y hääldatakse ühe häälikuna, ◌́ = kõrge toon, ◌̀ = madal toon, ◌̂ = langev toon, ˀ◌ = preglotaliseeritud häälik (eelneb lühike õhuvoolu katkestus kurgus), ◌’ = ejektiiv (konsonant, mille hääldamisel kurk korraks sulgub), ◌̥ = helitu häälik, ◌̃ = nasaliseerunud häälik (nina kaudu hääldatav häälik), ◌̰ = kärisev hääl (madala, ragiseva kõlaga),

$^{n}\circ$ tähendab, et enne kaashäälikut läheb osa õhust läbi nina, $\circ^{h}$ = aspireeritud konsonant (häälduse kaasneb lisahõngus), $\circ^{w}$ = labialiseeritud kaashäälik (hääldatakse ümarate huultega), $\circ^{j}$ = palataliseeritud (pehmendatud) häälik (hääldamisel liigub keel kõva suulae poole). **ɑ**, **æ**, **ɛ**, **ɪ**, **i̇**, **ɔ**, **ʊ**, **ʉ̈**, **ə**, **ʌ**, **ɒ**, **ɵ**, **y**, **ɵ**, **ø** on täishäälikud. Muud erimärgid tähistavad kaashäälikuid.

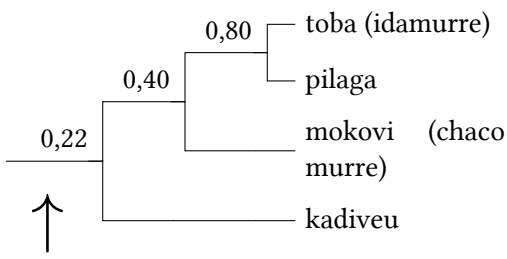
⚠ Mõne ülesandes kasutatud keele oskamine ei anna ülesande lahendamiseks mingeid eeliseid.

I osa. Guaikuru keelkond (Argentina, Brasiilia, Paraguay)

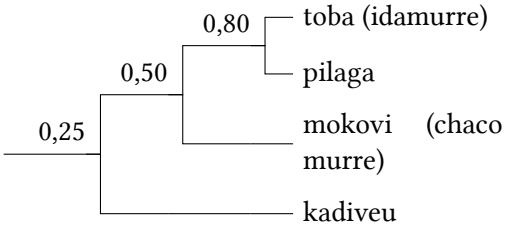
	toba (idamurre)	pilaga	mokovi (chaco murre)	kadiveu
pilv	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
tuli	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=ol:edi$_2$
kala	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nij:oɢo-d͡ʒegi$_3$
pea	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
tapma	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
kuu (taevakeha)	ʔawoʁojk$_1$	ʔaˈwoʃojk$_1$	ʃiɾajɣo$_2$	ep:enaj$_3$
nina	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
sool	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jok:i$_{-1}$
kivi	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
keel	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʃat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

käsitsi

A-algoritm



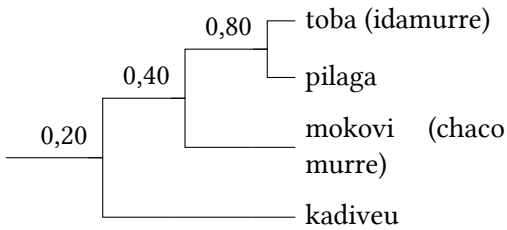
B-algoritm

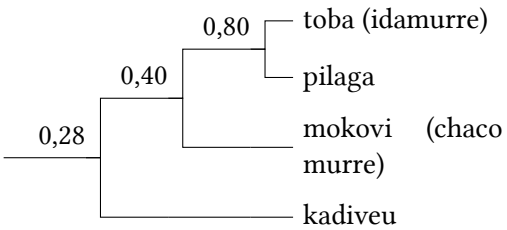


Stabiilsusindeksid:

pilv	0,50
tuli	0,50
kala	0,50
pea	0,75
tapma	1,00
kuu (taevakeha)	0,50
nina	1,00
sool	0,67
kivi	0,75
keel	0,50

automatiseeritud



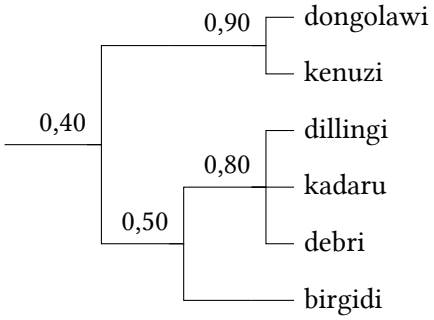
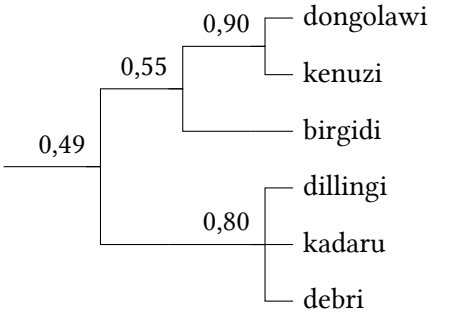
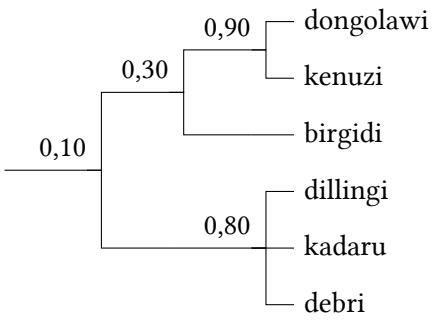
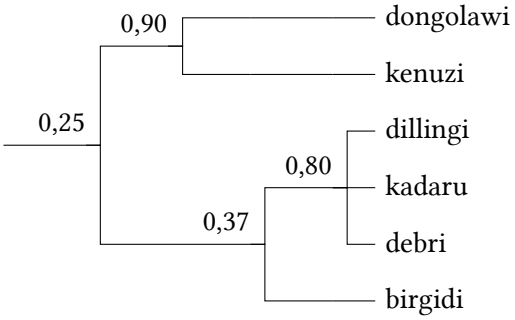


Stabiilsusindeksid:

pilv	0,50
tuli	0,50
kala	0,75
pea	0,75
tapma	1,00
kuu (taevakeha)	0,50
nina	1,00
sool	0,25
kivi	0,75
keel	0,50

II osa. Nuubia keelkond (Egiptus, Sudaan)

	dongolawi	kenuzi	dillingi	kadaru	debri	birgidi
tapma	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	fila:l-e$_1$
kuu (taevakeha)	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	ma:l$_2$
vesi	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
andma	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
hea	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
tuul	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
juuksed	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
kõht	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
magama	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
päike	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	A-algoritm	B-algoritm	
käsitsi			Stabiilsusindeksid: tapma 0,50 kuu (taevakeha) 0,83 vesi 1,00 andma 1,00 hea 0,50 tuul 0,50 juuksed 0,83 kõht 0,83 magama 0,83 päike 0,50
automatiseeritud			Stabiilsusindeksid: tapma 0,33 kuu (taevakeha) 0,50 vesi 0,50 andma 0,67 hea 0,50 tuul 0,50 juuksed 0,83 kõht 1,00 magama 0,50 päike 0,50

(A) **(2 punkti)** Kaashäälik **ʁ** häälдub nagu prantsuse *r* ehk keele tagaosas. Millisesse Dolgopolski klassi see kuulub ja kuidas te teada saite?

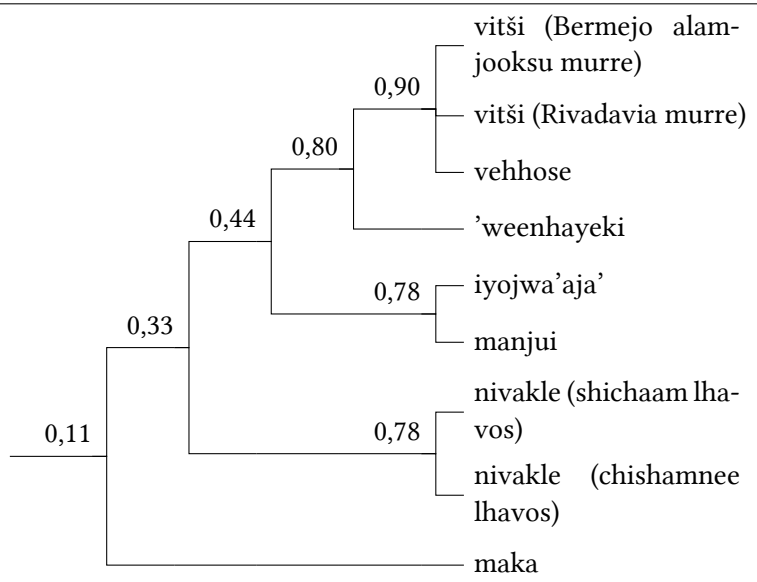
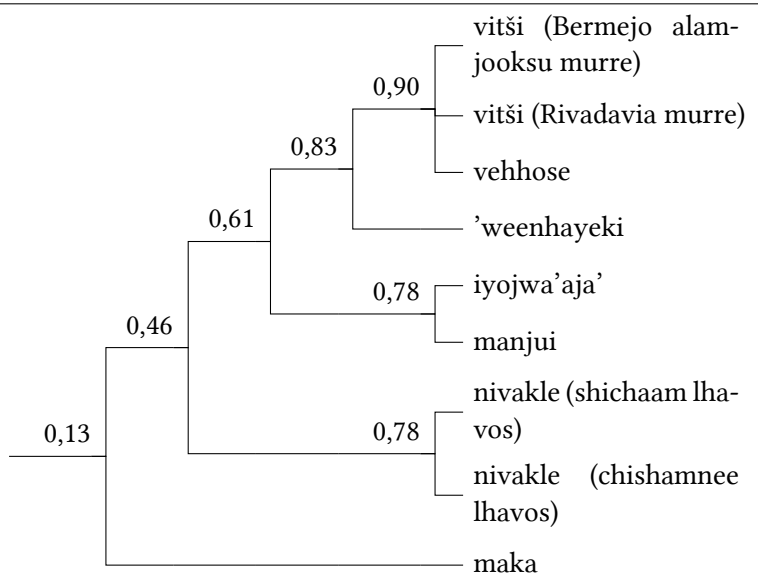
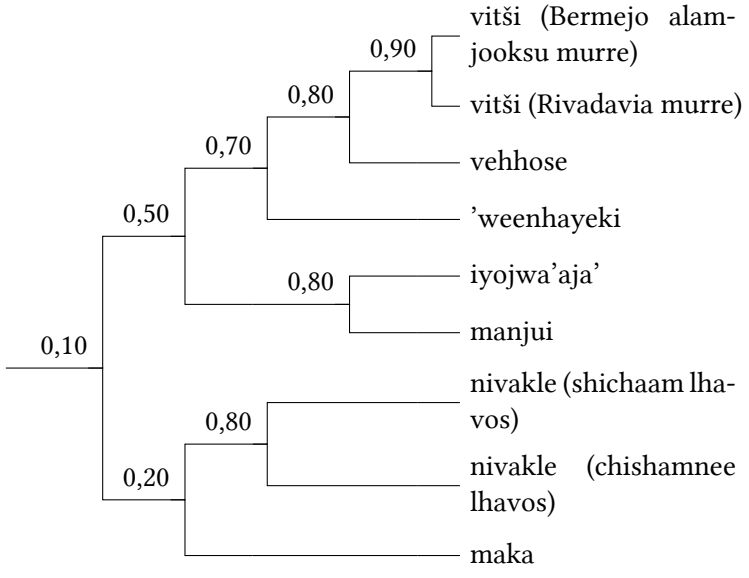
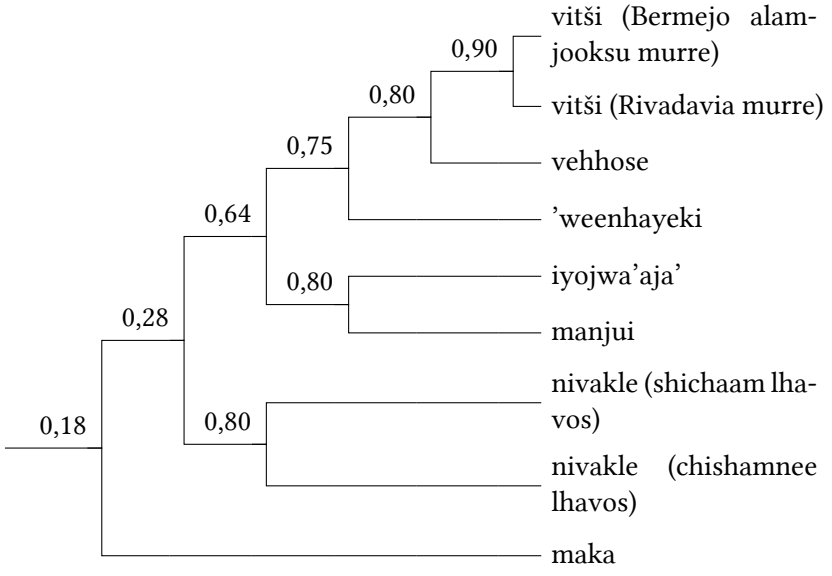
(B) **(2 punkti)** Nuubia keelte puu üleval vasakul on ainult üks kahest võimalikust puust selle algoritmi ja märkustetüübi kombinatsiooni puhul. Joonistage teine võimalik puu.

(C) **(2 punkti)** Nuubia keelte puu all vasakul on ainult üks kahest võimalikust puust selle algoritmi ja märkustetüübi kombinatsiooni puhul. Joonistage teine võimalik puu.

(D) **(2 punkti)** Leksikostatistiline kaugus 0,49 (mis on määratud Nubia puu juurele üleval paremal) on ümardatud kahe kümnendkohani, nagu ka mõned teised selle ülesande kaugused. Mis on täpne kaugus?

III osa. Mataguai keelkond (Argentina, Boliivia, Paraguay)

	vitši (Bermejo alamjooksu murre)	vitši (Rivadavia murre)	vehhose	’weenhayeki	iyojwa’aja’	manjui	nivakle (shichaam lhavos)	nivakle (chishamnee lhavos)	maka
tuli	ʔitoχ$_{1}$	ʔitɔx$_{1}$	ʔitɑh$_{1}$	ʔi:tɑχ$_{1}$	ˈh͡wat$_{2}$	ˈʔe͡itʲe$_{1}$	ʔitɑx$_{1}$	ʔitɑx$_{1}$	feʔt$_{2}$
kala	ʔwahat$_{1}$	wahat$_{1}$	wahat$_{1}$	ʔwa:hat$_{1}$	siˈʔjus$_{-1}$	ʃiˈʔjus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	sehet͡s$_{-1}$
jalg	=pat͡ʃu$_{1}$	=qɔlɔ$_{2}$	=pat͡ʃo$_{1}$, =kɑla$_{2}$	=pa:kʲ’oʔ$_{1}$	=ˈsat$_{3}$	=kaˈʔlaʔ$_{2}$	=ɸoʔ$_{4}$	=ɸoʔ$_{4}$	=f’iʔ$_{5}$
vesi	ʔinot$_{1}$	ʔinɔt$_{1}$	wah$_{2}$	ʔinɑ:t$_{1}$	ʔiˈn̪at$_{1}$	ʔaˈʔnat$_{1}$	jinɑʔt$_{1}$	jinɑʔt$_{1}$	iweliʔ$_{3}$
andma	=ʔwen̪-u$_{1}$	=wen̪-u$_{1}$	=ʔwen̪-o$_{1}$	=ʔwen̪-oʔ$_{1}$	=ˈwɛh͡n-aʔm$_{2}$	=ˈhajʔ$_{3}$, =ˈwɛn$_{2}$	=xut$_{4}$	=xut-ej$_{4}$	tis-ix$_{5}$
hea	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ˈʔes$_{1}$	ˈʔe͡is$_{1}$	ʔis$_{1}$	ʔis$_{1}$	t=ejk’un-ej$_{2}$
tuul	ʔinwokʷ$_{1}$	ʔinwɔk$_{1}$	ʔihwokʷ$_{1}$	=ja:ɬ$_{2}$, =xʷoxʷ$_{3}$	ˈh͡lah͡wuʔ$_{4}$	ˈh͡lah͡wuuʔ$_{4}$	ɬaβiʔm$_{5}$	ɬaβiʔm$_{5}$	t’unik’i$_{6}$
puu	haʔlo$_{1}$	halɔ$_{1}$	haʔlɑ$_{1}$	haʔlɑʔ$_{1}$	ʔaˈʔlaʔ$_{1}$	ʔaˈʔla-k$_{1}$	ʔaʔkxi-juk$_{2}$	jiʔk͡lɑʔ$_{1}$	naxka-k$_{3}$
juuksed	=ʔwule-j$_{1}$	=wule-j$_{1}$	=ʔwole-j$_{1}$	=ʔwo:le-ç$_{1}$, hi:lenaχ$_{2}$	=ˈʔwole$_{1}$	=ˈʔwole-j$_{1}$	=ʃateʔt͡ʃ$_{3}$	=jeʔs$_{4}$	=ʔewkux-it͡s$_{5}$
tapma	=lon$_{1}$	=lɔn$_{1}$	=lɑn$_{1}$	=lɑ:n̪$_{1}$	=ˈlaʔan$_{1}$	=ˈlan$_{1}$	=k͡lɑn$_{1}$	=k͡lɑn$_{1}$	=lan$_{1}$

	A-algoritm	B-algoritm	
käsitsi			Stabiilsusindeksid: tuli 0,78 kala 1,00 jalg 0,33 vesi 0,78 andma 0,44 hea 0,89 tuul 0,33 puu 0,78 juuksed 0,67 tapma 1,00
automatiseeritud			Stabiilsusindeksid: tuli 0,78 kala 0,44 jalg 0,33 vesi 0,56 andma 0,67 hea 0,89 tuul 0,22 puu 0,67 juuksed 0,67 tapma 1,00

IV osa. Mongoli keelkond (Hiina Rahvavabariik, Mongoolia, Venemaa)

(E) (10 punkti) Vaadake järgnevat sõnaloetelu. Arvutage stabiilsusindeks,nii käsitsi tehtud kui ka automatiseeritud märkuste jaoks.

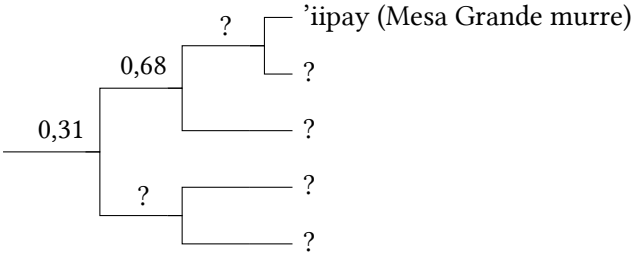
Et teid pisut aidata, oleme sõna ‘kõik’ jaoks mõlemad stabiilsusindeksid juba välja arvutanud. Need on – suvalises järjekorras – 0,36 ja 0,40.

	daguuri (hailari muure)	hamnigani (mandžuu-ria murre)	burjaadi (hori murre)	uusbarga	ööldi	hošuudi	kalmõki	halha	ordose	šira-juguri	bonani
kõik	hɔː$_{1}$	bʊlt$_{2}$	bʊxɨː$_{3}$	bygd$_{4}$	t͡sug$_{5}$	lug$_{5}$	t͡suk$_{5}$, xamak$_{-1}$	pux$_{3}$, pugt$_{4}$, xamăɢ$_{-1}$	pyɣyte$_{4}$, xamuk^{h}$_{-1}$	t͡ʃ^{h}uq$_{5}$	hanə$_{-2}$
puukoor	hai̯l̩s$_{1}$	qalihʊn$_{1}$	χɔltəhɔn$_{2}$	xal^{j}huː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xɔɮt^{h}ɔ̆s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
kõht	keːli$_{1}$	gətəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
lind	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡ʑer$_{2}$
tuli	ɢali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
rada	terg-uːl$_{1}$	qarɢʊ̯i$_{2}$	χarɢɨ$_{2}$, zam$_{-1}$	zam$_{-1}$	d͡zam$_{-1}$	d͡zam$_{-1}$	xaː-lʁə$_{3}$	t͡sam$_{-1}$	t͡ʃam$_{-1}$	mør$_{4}$	mor$_{4}$
sool	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
ujuma	unpa-du$_{1}$	ʊmba-$_{1}$	t^{h}amar-$_{2}$	umb-$_{1}$	sele-$_{3}$	umba-$_{1}$	us-t͡ɕi-$_{4}$, øːm-$_{5}$	siɮĭ-$_{3}$	usu-t͡ʃ^{h}i-la-$_{4}$	umpa-$_{1}$	mba-$_{1}$
vesi	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
tuul	kei̯n$_{1}$	halkin$_{2}$	halxin$_{2}$	halxi$_{2}$	sal^{j}xin$_{2}$	salk^{j}i$_{2}$	sal^{j}kn$_{2}$	saɮxĭ$_{2}$	k^{h}iː$_{1}$	k^{h}iː$_{1}$	ki$_{1}$

V osa. Juma keelkond (Mehhiko, Ameerika Ühendriigid)

(F) **(8 punkti)** Vaadake järgnevat sõnaloetelu. All näete puud, mis konstrueeriti sama loetelu alusel. Mõningad andmed (keelte nimed ja leksikostatistilised kaugused) on puudu. Täitke lüngad. Pange kirja, kas puu on käsitsi tehtud või automaatne, samuti, kas see on tehtud algoritmi A või B järgi.

	mohave	kokopa	javapai	tiipai (jamuli)	’iipay (Mesa Grande murre)
lühike	wena=wen-a_1	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i$_2$	ləˈʔuɲ$_1$	mə=put-k$_3$
lind	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔ=ˈʃa$_2$	ʔaː=ʂaː$_2$
luu	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a$_1$	ˈak$_1$	aq$_1$
kuiv	i=roː-v-k$_1$	ˈʂ=ʔar$_2$	ˈru-β-i$_1$	ˈs=ʔaːj$_3$	saːj$_3$
liha	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeː=ˈθo-β-a$_3$	ˈkʷak$_4$	kukʷaːj-p$_1$
kael	maʌaqe$_1$	ˈm=puɬ$_2$	ˈmlqí$_1$	iː=ˈpuɬ$_2$	iː=puɬ$_2$
nägema	i=juː-k$_1$	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
saba	iː=ʔar$_1$	ˈʃ=juʂ$_2$	ˈβ=hé$_3$	ʃə=ˈjuɬ$_2$	xə=juɬ$_2$
kaks	havik-k$_1$	ˈx=wak$_1$	ˈh͡wâk-i$_1$	xə=ˈwak$_1$	xə=wak$_1$
aasta	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a$_3$	mat-ˈwam$_2$	ʔaːnʲ$_{-1}$



(G) **(20 punkti)** Juma keelte jaoks on loodud veel mõned puud, kus puude juurte vahel on järgmised leksikostatistilised kaugused (iga puu vasakpoolses otsas):

1. 0,20
2. 0,23
3. 0,24

Joonistage neist igaühe kohta puu. Pange iga puu kohta kirja, kas see on käsitsi tehtud või automaatne, samuti, kas see on tehtud algoritmi A või B järgi.

(H) **(3 punkti)** Kaks osaülesandes **(G)** esitatud kaugust on ümardatud kahe kümnendkohani: 0,23 on ümardatud 0,225 pealt. Milline teine kaugus on ümardatud ning mis on selle täpne väärtus?

(I) **(4 punkti)** Selgitage, kuidas arvutatakse stabiilsusindeksid.

(J) **(5 punkti)** Selgitage, kuidas arvutatakse leksikostatistilised kaugused.

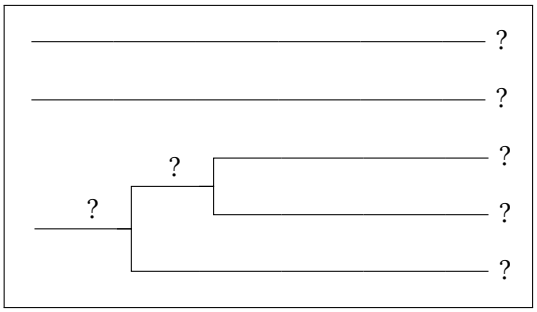
(K) **(4 punkti)** Selgitage A- ja B-algoritmide vahelist erinevust.

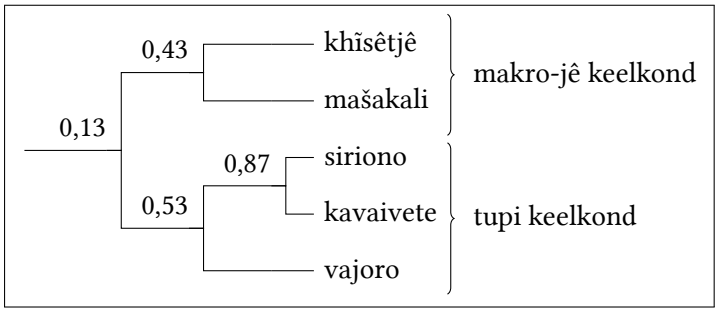
VI osa. Makro-jê keelkond ja tupi keelkond (Brasiilia, Boliivia)

(L) **(28 punkti)** Makro-jê ja tupi on kaks suuremat keelkonda Lõuna-Ameerikas. Mõned keeleteadlased usuvad, et nad on omavahel kaugemalt suguluses. Vaadake järgnevaid sõnaloetelusid.

	A	B	Γ	Δ	E
puukoor	e=ˈe-ke	hʷĩ=ˈkʰʌ	kʉp=ˈpɛ	mĩβ͡m=ˈt͡ɕaj	=ˈpɛ
kõht	ˈe=ɾje	=ˈtʰigi	=ãˈɯ̃n	=ˈtæj	=ɾɛˈwɛk
veri	e=ˈɾuki	=kaˈⁿbɽo	=d͡ʒ=aˈʉ	=ˈhɛβ͡p	=ɾuˈɨ
põletama	=ˈɾa͡i	=ɾɔ=ˈkʰɻə̃	=poˈkʷa	mũ=...=ˈha β͡p	=ɾaˈpɨ
rasv	e=ˈkiɾa	=ˈtʰwəmɨ	=ˈd͡ʒ=ap	=ˈtuβ͡p	=ˈkap
jalg	ˈe=i	=ˈhʷaji	=ˈβi	=pɒˈta	=ˈpɨ
käsi	ˈe=o	=ɲĩˈkʰɻa	=ˈβo	=ˈɲĩβ͡m	=ˈpɔ
raske	e=ˈusi	=wɨˈtʰĩ	=poˈti	=β͡pˈtəj	=pɔˈɨj
maks	ˈe=ja	=ˈⁿba	=piˈa	=t͡ɕɨβ͡pkĩˈnɑ̃j	=pɨˈʔa
uus	e=ˈjasu	=ˈⁿdɨwɨ	=paˈgop	=ˈtɨβ͡p	=pɨaˈu
juur	e=ˈɾao	=jaˈɽe	kʉp=kʉjoˈpɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕaˈti͡ə	=ɾaˈpɔ
nahk	ˈe=ɨ	=ˈkʰʌ	=ˈpɛ	=ˈt͡ɕaj	=ˈpit
saba	e=ˈɾoko͡i	=ˈⁿbɨ	=d͡ʒ=oˈkʷaj	=nãː=ˈkɨβ͡p	=ˈɾaj
valge	ˈe=ʃɨ	=jaˈkʰa	=d͡ʒiˈɾa	=β͡pˈdouɻ	=ˈsĩŋ
tiib	e=ˈheo	=jaˈɽa	=pɛˈo	=ɲĩˈmãuɻ	=pɛˈpɔ, =jɨˈwa

All näete kahte puud, mis konstrueeriti samade loetelude alusel. Mõningad andmed (keelte nimed ja leksikostatistilised kaugused) on puudu. Täitke lüngad. Pange iga puu kohta kirja, kas see on käsitsi tehtud või automaatne, samuti, kas see on tehtud algoritmi A või B järgi.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ Käsitsi tehtud märkused ning stabiilsusindeksid on siin osaülesandes meelega ära jäetud.

(M) (10 punkti) Dolgopolski klassidel põhinevad automaatsed protsessid võivad viia valede tulemusteni. Selles näites tuvastab automaatne protsess rohkem sarnasusi siriono ja ühe teatud makro-jê keele (khĩsêtjê keele) vahel kui siriono ja teiste tupi keelte vahel. Tehke ettepanek modifitseeritud automaatse protsessi kohta, mis annaks makro-jê ja tupi sõnastike puhul õige klassifikatsiooni, ja kirjeldage seda *lühidalt*.

⚠ Seda osaülesannet hinnatakse ainult sel juhul, kui parimate tulemustega võistkonnad lõpetavad viigiga.

Autorid tänavad Alejandro Vidali, Maria Konošenkot, Ilja Gruntovit ja Jamthô Suyat, kes lahkelt vastasid küsimustele üksikute keelte kohta. —*Andrei Nikulin, Milena Veneva*

Toimetajad: Ivan Deržanski (tehn. toim.), Hugh Dobbs, Stanislav Gurevitš, Boriss Iomdin, Liam McKnight, Andrei Nikulin (vast. toim.), Aleksejs Peguševs, Jan Petr, Aleksandr Piperski, Marija Rubinštein, Milena Veneva, Elysia Warner.

Eesti tekst: Axel Jagau.

Edu!