A douăzeci și una Olimpiadă internațională de lingvistică

Brasília (Brazilia), 23–31 iulie 2024

Problema pentru competiția pe echipe

Lexicostatistica este un grup de metode create pentru a estima cât de apropiate sunt diferite limbi pe baza vocabularului lor. Aceste metode se aplică în mod normal la liste lungi de cuvinte care au fost adnotate manual de către experți, care indică dacă consideră că o anumită pereche de cuvinte provine din aceeași sursă. Totuși, uneori lingviștii aplică metode lexicostatistice și la liste de cuvinte care au fost adnotate prin proceduri automate. O astfel de procedură se bazează pe conceptul de *clase consonantice*, introdus de lingvistul sovieto-israelit Aharon Dolgopolsky în 1964.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ʯ	Y.	j ç (la începutul rădăcinii)	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 (la începutul rădăcinii)	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, vocale și j ç w ʍ (exceptând la începutul rădăcinii)						

Clasele consonantice ale lui Dolgopolsky

Mai jos veți găsi adnotate fragmente ale unor liste de cuvinte din mai multe familii de limbi ale lumii. Adnotările sunt date ca cifre în indice. Pe baza acestor liste, au fost construiți arbori genealogici ai familiilor de limbi utilizând două versiuni simplificate ale așa-numitului algoritm *StarlingNJ*, iar un *indice de stabilitate* a fost atribuit fiecărui cuvânt. Arborii și indicii de stabilitate de mai sus se bazează pe liste de cuvinte adnotate manual, pe când cei de mai jos se bazează pe liste adnotate automat. Doi arbori au fost construiți pentru fiecare listă de cuvinte, utilizând două versiuni diferite ale algoritmului: Algoritmul A și Algoritmul B. N.B.: în anumite cazuri, există mai mulți arbori posibili pentru o singură listă de cuvinte; în acele cazuri, un singur arbore a fost ales în mod aleatoriu. Fiecărui nod din fiecare arbore i se atribuie o distanță lexicostatistică. Cu cât distanța este mai mare, cu atât mai apropiate sunt limbile. Un termen mai precis ar fi deci „distanță lexicostatistică inversată” în loc de „distanță lexicostatistică”. Pentru simplificare, termenul „distanță lexicostatistică” este utilizat în problemă.

Indicii de stabilitate și distanțele lexicostatistice sunt rotunjite la două zecimale. Dacă a treia zecimală este mai mică de 5, se rotunjește la valoarea inferioară; altfel, se rotunjește la valoarea superioară. Spre exemplu, 2,836 este rotunjit la 2,84, 0,705 la 0,71, iar 0,703 la 0,70. Rotunjirea se aplică doar la valorile prezentate cititorilor umani. Cu alte cuvinte, computerul care rulează algoritmul „vede” valorile nerotunjite.

N.B.: despre unele cuvinte se știe sau se bănuiește că au fost împrumutate din alte limbi. Spre exemplu, cuvântul **jokːi** ‘sare’ din limba kadiwéu este împrumutat din cuvântul guarani **jukɨ**, iar **ʔaːnʲ** ‘an’ din ipai (de Mesa Grande) este împrumutat din cuvântul spaniol **ˈaɲo**.

În unele cazuri, se dau mai multe sinonime cu același sens, separate prin virgulă. Un exemplu este ‘picior’ în limba vejoz.

În datele de mai jos, toate prefixele se separă prin semnul „=”, iar toate sufixele se separă prin semnul „-”. Unele cuvinte nu se folosesc decât cu prefixe. Acele cuvinte încep cu semnul „=”.

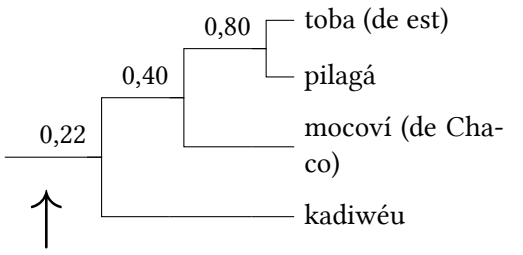
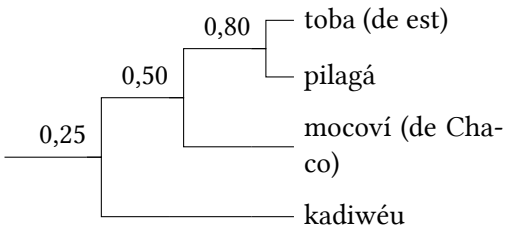
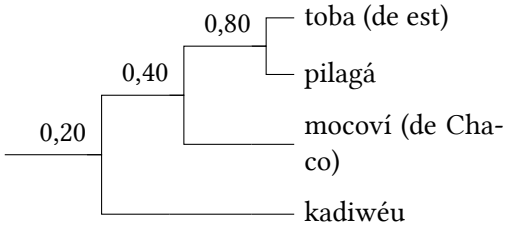
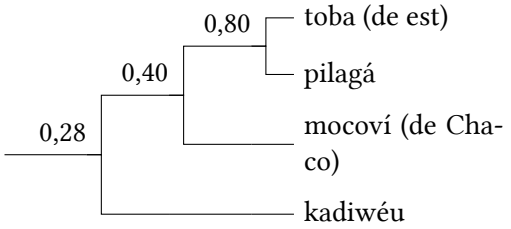
Datele au fost transcrise în Alfabetul Fonetic Internațional. ˈ = accent primar, ˌ = accent secundar (mai slab ca accentul primar), ◌ː = sunet lung, ◌̆ = sunet foarte scurt, $\widehat{XY}$ = X și Y se pronunță ca un singur sunet, ◌́ = ton înalt, ◌̀ = ton grav, ◌̂ = ton descendent, ◌ˀ = sunet preglotalizat (precedat de blocarea scurtă a fluxului de aer în gât), ◌’ = sunet ejectiv (pronunțat prin blocarea scurtă a fluxului de aer în gât), ◌̥ = sunet surd, ◌̃ = sunet nazalizat (pronunțat prin nas), ◌̰ = laringalizare (sunet grav,

zgârietor), ⁿ◌ arată că aceste consoane sunt precedate de aer prin cavitatea nazală, ◌ʰ = consoană aspirată (pronunțată cu un suflu de aer), ◌ʷ = consoană labializată (pronunțată cu buzele rotunjite), ◌ʲ = sunet palatalizat (pronunțat cu o parte a limbii apropriindu-se de palatul dur). **ɑ**, **æ**, **ɛ**, **ɪ**, **ɨ**, **ɔ**, **ʊ**, **ʉ**, **ə**, **ʌ**, **ɒ**, **ɘ**, **ɣ**, **ɵ**, **ø** sunt vocale. Celelalte caractere speciale sunt consoane.

⚠ Cunoașterea limbilor menționate în problemă nu oferă niciun avantaj pentru rezolvarea problemei.

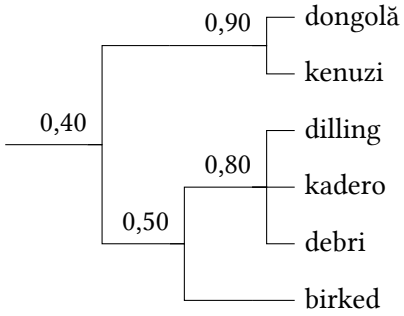
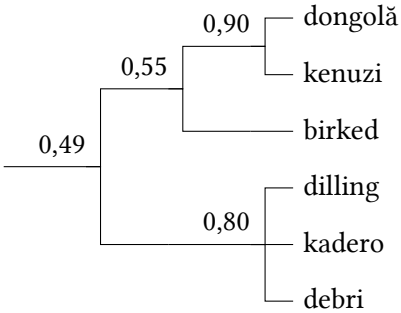
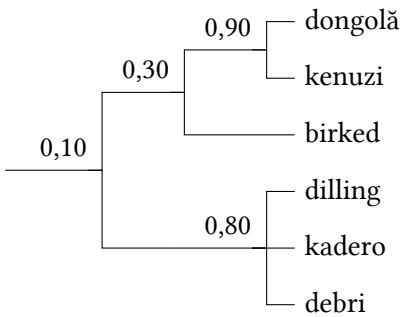
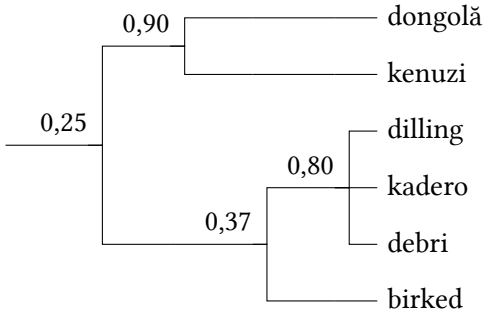
Partea I. Familia guaicuru (Argentina, Brazilia, Paraguay)

	toba (de est)	pilagá	mocoví (de Chaco)	kadiwéu
nor	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
foc	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=ol:edi$_2$
pește	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nij:oɢo-d͡ʒegi$_3$
cap	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
a ucide	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
lună	ʔawoʁojk$_1$	ʔaˈwoʕojk$_1$	ʃiɾajɣo$_2$	ep:enaj$_3$
nas	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
sare	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jok:i-$_1$
piatră	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
limbă	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʕat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

	Algoritmul A	Algoritmul B	
manual			Indici de stabilitate: nor 0,50 foc 0,50 pește 0,50 cap 0,75 a ucide 1,00 lună 0,50 nas 1,00 sare 0,67 piatră 0,75 limbă 0,50
automat			Indici de stabilitate: nor 0,50 foc 0,50 pește 0,75 cap 0,75 a ucide 1,00 lună 0,50 nas 1,00 sare 0,25 piatră 0,75 limbă 0,50

Partea a II-a. Familia nubiană (Egipt, Sudan)

	dongolă	kenuzi	dilling	kadero	debri	birked
a ucide	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	fila:l-e$_1$
lună	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	ma:l$_2$
apă	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
a da	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
bun	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
vânt	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
păr	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
burtă	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
a dormi	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
soare	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	Algoritmul A	Algoritmul B	
manual			Indici de stabilitate: a ucide 0,50 lună 0,83 apă 1,00 a da 1,00 bun 0,50 vânt 0,50 păr 0,83 burtă 0,83 a dormi 0,83 soare 0,50
automat			Indici de stabilitate: a ucide 0,33 lună 0,50 apă 0,50 a da 0,67 bun 0,50 vânt 0,50 păr 0,83 burtă 1,00 a dormi 0,50 soare 0,50

(A) **(2 puncte)** Consoana **ʁ** se pronunță precum *r* în franceză, cu partea din spate a limbii. Cărei clase Dolgopolsky aparține și cum v-ați dat seama?

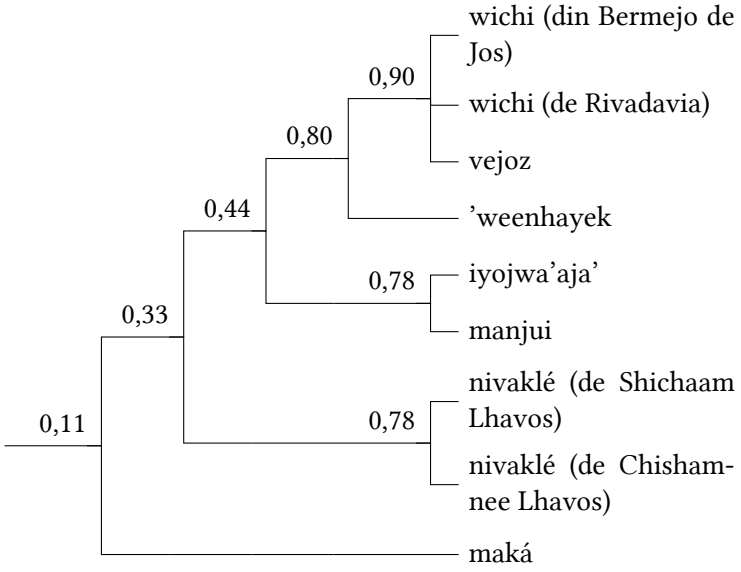
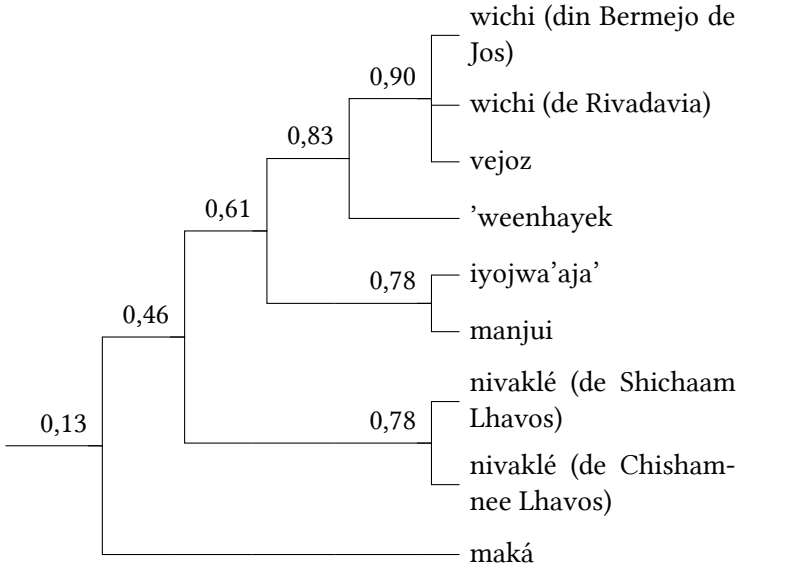
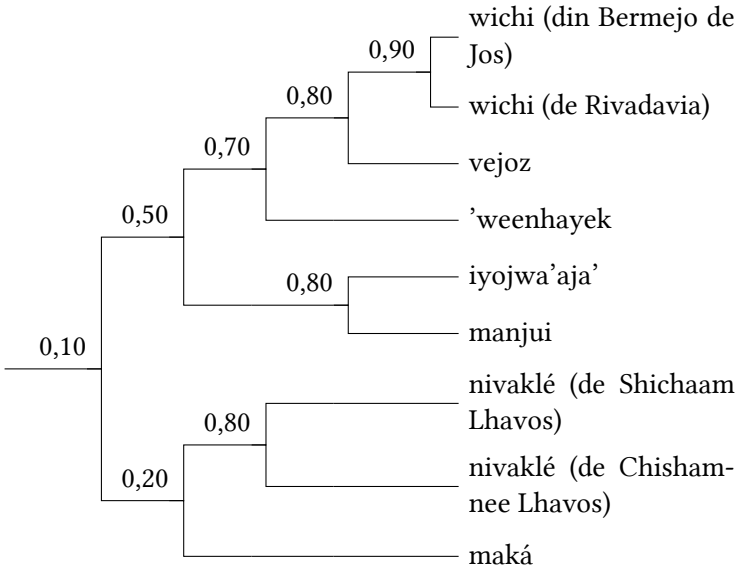
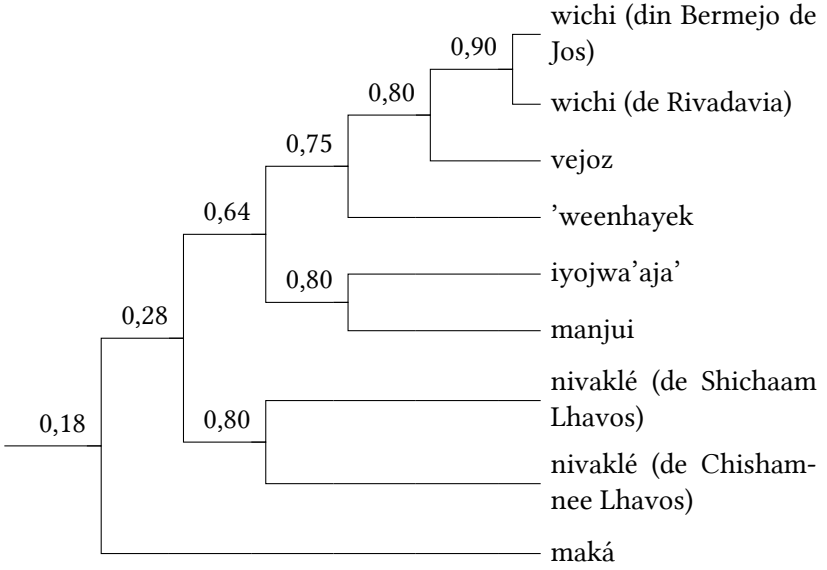
(B) **(2 puncte)** Arborele nubian din stânga sus este doar unul din doi arbori posibili pentru această combinație de algoritm și tip de adnotare. Desenați celălalt arbore posibil.

(C) **(2 puncte)** Arborele nubian din stânga jos este doar unul din doi arbori posibili pentru această combinație de algoritm și tip de adnotare. Desenați celălalt arbore posibil.

(D) **(2 puncte)** Distanța lexicostatistică 0,49 (atribuită rădăcinii arborelui nubian din dreapta sus) a fost rotunjită la două zecimale, precum alte distanțe din problemă. Care este distanța exactă?

Partea a III-a. Familia mataco (Argentina, Bolivia, Paraguay)

	wichi (din Bermejo de Jos)	wichi (de Rivada-via)	vejoz	’weenhayek	iyojwa’aja’	manjui	nivaklé (de Shichaam Lhavos)	nivaklé (de Chisham-nee Lhavos)	maká
foc	ʔitoχ$_{1}$	ʔitɔx$_{1}$	ʔitɑh$_{1}$	ʔiːtɑχ$_{1}$	ˈh͡wat$_{2}$	ˈʔe͡itʲe$_{1}$	ʔitɑx$_{1}$	ʔitɑx$_{1}$	feʔt$_{2}$
pește	ʔwahat$_{1}$	wahat$_{1}$	wahat$_{1}$	ʔwaːhat$_{1}$	siˈʔjus$_{-1}$	ʃiˈʔjus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	seheț$_{-1}$
picior	=pat͡ʃu$_{1}$	=qɔlɔ$_{2}$	=pat͡ʃo$_{1}$, =kɑla$_{2}$	=paːkʲ’oʔ$_{1}$	=ˈsat$_{3}$	=kaˈʔlaʔ$_{2}$	=ɸoʔ$_{4}$	=ɸoʔ$_{4}$	=fʼiʔ$_{5}$
apă	ʔinot$_{1}$	ʔinɔt$_{1}$	wah$_{2}$	ʔinɑːt$_{1}$	ʔiˈnʲat$_{1}$	ʔaˈʔnat$_{1}$	jinɑʔt$_{1}$	jinɑʔt$_{1}$	iweliʔ$_{3}$
a da	=ʔwen̥-u$_{1}$	=wen̥-u$_{1}$	=ʔwen̥-o$_{1}$	=ʔwen̥-oʔ$_{1}$	=ˈwɛh͡n-aʔm$_{2}$	=ˈhajʔ$_{3}$, =ˈwɛn$_{2}$	=xut$_{4}$	=xut-ej$_{4}$	tis-ix$_{5}$
bun	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ˈʔes$_{1}$	ˈʔe͡is$_{1}$	ʔis$_{1}$	ʔis$_{1}$	t=ejkʼun-ej$_{2}$
vânt	ʔinwokʷ$_{1}$	ʔinwɔk$_{1}$	ʔihwokʷ$_{1}$	=jaːɬ$_{2}$, =xʷoxʷ$_{3}$	ˈh͡lah͡wuʔ$_{4}$	ˈh͡lah͡wuuʔ$_{4}$	ɬaβiʔm$_{5}$	ɬaβiʔm$_{5}$	tʼunikʼi$_{6}$
copac	haʔlo$_{1}$	halɔ$_{1}$	haʔlɑ$_{1}$	haʔlɑʔ$_{1}$	ʔaˈʔlaʔ$_{1}$	ʔaʔla-k$_{1}$	ʔaʔkxi-juk$_{2}$	jiʔk͡lɑʔ$_{1}$	naxka-k$_{3}$
păr	=ʔwule-j$_{1}$	=wule-j$_{1}$	=ʔwole-j$_{1}$	=ʔwoːle-ç$_{1}$, hiːlenaχ$_{2}$	=ˈʔwole$_{1}$	=ˈʔwole-j$_{1}$	=ʃateʔt͡ʃ$_{3}$	=jeʔs$_{4}$	=ʔewkux-it͡s$_{5}$
a ucide	=lon$_{1}$	=lɔn$_{1}$	=lɑn$_{1}$	=lɑːn̥$_{1}$	=ˈlaʔan$_{1}$	=ˈlan$_{1}$	=k͡lɑn$_{1}$	=k͡lɑn$_{1}$	=lan$_{1}$

<table>
<tr><th></th><th>Algoritmul A</th><th>Algoritmul B</th><th></th></tr>
<tr><td>manual</td><td> 
</td><td> 
</td><td>Indici de stabilitate:
foc 0,78
pește 1,00
picior 0,33
apă 0,78
a da 0,44
bun 0,89
vânt 0,33
copac 0,78
păr 0,67
a ucide 1,00</td></tr>
<tr><td>automat</td><td> 
</td><td> 
</td><td>Indici de stabilitate:
foc 0,78
pește 0,44
picior 0,33
apă 0,56
a da 0,67
bun 0,89
vânt 0,22
copac 0,67
păr 0,67
a ucide 1,00</td></tr>
</table>

Partea a IV-a. Familia mongolică (Republica Populară Chineză, Mongolia, Rusia)

(E) (10 puncte) Examinați următoarea listă de cuvinte. Calculați indicii de stabilitate corespunzători adnotărilor manuale și automate.

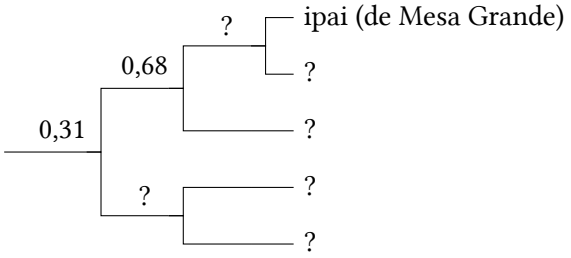
Pentru a vă ajuta, am calculat deja ambii indici de stabilitate pentru cuvântul ‘toți’. În ordine aleatorie, aceștia sunt 0,36 și 0,40.

	dagură (de Hailar)	hamnigană (manchu)	buriată (de Khori)	bargu nouă	ööld	khoșută	kalmâkă	khalkha	ordos	yugura de est	bonan
toți	hɔː$_{1}$	bʊlt$_{2}$	bʊxɨː$_{3}$	bygd$_{4}$	t͡sug$_{5}$	lug$_{5}$	t͡suk$_{5}$, xamak$_{-1}$	pux$_{3}$, pugt$_{4}$, xamaˇɢ$_{-1}$	pyɣyte$_{4}$, xamuk^{h}$_{-1}$	t͡ʃ^{h}uq$_{5}$	hanə$_{-2}$
scoarță	hai̯l͡s$_{1}$	qalihʊn$_{1}$	χɔltəhɔn$_{2}$	xaʲhuː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xɔɮt^{h}ɔ̆s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
burtă	keːli$_{1}$	gʊtəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
pasăre	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡ʑer$_{2}$
foc	gali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
drum	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$_{4}$
sare	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
a înota	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-$_{1}$
apă	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
vânt	kei̯n$_{1}$	halkin$_{2}$	halxin$_{2}$	halxi$_{2}$	salʲxin$_{2}$	salkʲi$_{2}$	salʲkn$_{2}$	saɮxĭ$_{2}$	k^{h}iː$_{1}$	k^{h}iː$_{1}$	ki$_{1}$

Partea a V-a. Familia yuman (Mexic, Statele Unite ale Americii)

(F) **(8 puncte)** Examinați următoarea listă de cuvinte. Mai jos se dă un arbore construit pe baza aceleiași liste. Unele date (numele limbilor și distanțele lexicostatistice) lipsesc. Completați spațiile libere. Precizați dacă arborele este manual sau automat, precum și dacă a fost generat utilizând algoritmul A sau B.

	mohave	cocopa	yavapai	tiipai (de Jamul)	ipai (de Mesa Grande)
scurt	wena=wen-a_1	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i_2	ləˈʔuɲ$_1$	mə=put-k_3
pasăre	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔˈʃa$_2$	ʔaː=ʂaː$_2$
os	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a_1	ˈak$_1$	aq$_1$
uscat	i=roː-v-k$_1$	ˈʂ=ʔar$_2$	ˈru-β-i$_1$	ˈs=ʔaːj$_3$	saːj$_3$
carne	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeːˈθo-β-a_3	ˈkʷak$_4$	kukʷaːj-p_1
gât	maʌaqe$_1$	ˈm=puk$_2$	ˈmlqí$_1$	iːˈpuk$_2$	iː=puk$_2$
a vedea	i=juː-k$_1$	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
coadă	iː=ʔar$_1$	ˈʃ=juʌ$_2$	ˈβ=hé$_3$	ʃəˈjuɬ$_2$	xə=juɬ$_2$
doi	havik-k_1	ˈx=wak$_1$	ˈh͡wâk-i_1	xəˈwak$_1$	xə=wak$_1$
an	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a_3	matˈwam$_2$	ʔaːnʲ$_{-1}$



(G) **(20 de puncte)** Au fost generați câțiva alți arbori pentru familia yuman, cu următoarele distanțe lexicostatistice la rădăcina arborelui (distanțele lexicostatistice din marginea stângă a fiecărui arbore):

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

Desenați fiecare din acești arbori. Pentru fiecare arbore, specificați dacă este manual sau automat, precum și dacă a fost generat utilizând algoritmul A sau B.

(H) **(3 puncte)** Două din distanțele date în cerința **(G)** au fost rotunjite la două zecimale: 0,23 a fost rotunjit de la 0,225. Ce altă distanță a fost rotunjită și care este valoarea ei exactă?

(I) **(4 puncte)** Explicați cum sunt calculați indicii de stabilitate.

(J) **(5 puncte)** Explicați cum sunt calculate distanțele lexicostatistice.

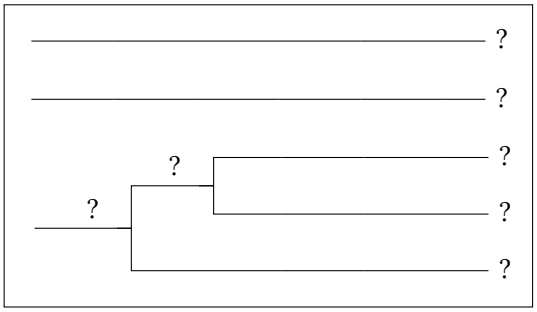
(K) **(4 puncte)** Explicați diferența dintre algoritmii A și B.

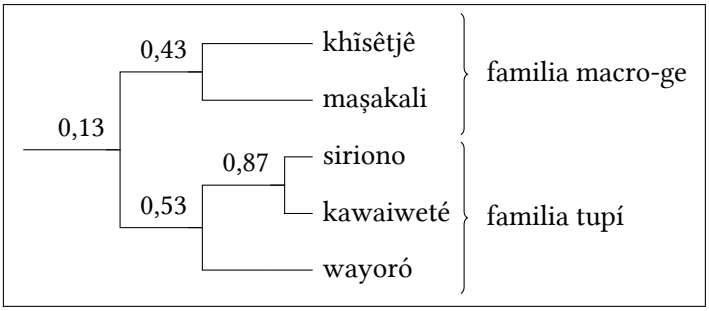
Partea a VI-a. Familia macro-ge și familia tupí (Brazilia, Bolivia)

(L) (28 de puncte) Macro-ge și tupí sunt două familii majore de limbi din America de Sud. Unii lingviști le consideră rude îndepărtate. Examinați următoarele liste de cuvinte.

	A	B	Γ	Δ	E
scoarță	e=ˈe-ke	hʷĩ=ˈkʰʌ	kʉp=ˈpɛ	mĩβ͡m=ˈt͡ɕaj	=ˈpɛ
burtă	ˈe=ɾje	=ˈʈʰigi	=ãˈʉ̃n	=ˈtæj	=ɾɛˈwɛk
sânge	e=ˈɾuki	=kaˈⁿbɻo	=d͡ʒ=aˈʉ	=ˈhɛβ͡p	=ɾuˈɨ
a arde	=ˈɾai̯	=ɾɔ=ˈkʰɹɜ̃	=poˈkʷa	mũ=...=ˈhaβ͡p	=ɾaˈpɨ
grăsime	e=ˈkiɾa	=ˈʈʰwɜmɨ	=ˈd͡ʒ=ap	=ˈtuβ͡p	=ˈkap
picior	ˈe=i	=ˈhʷaji	=ˈβi	=pɒˈta	=ˈpɨ
mână	ˈe=o	=ɲɨ̃ˈkʰɹa	=ˈβo	=ˈɲĩβ͡m	=ˈpɔ
greu	e=ˈusi	=wɨˈʈʰĩ	=poˈti	=β͡pˈtəj	=pɔˈɨj
ficat	ˈe=ja	=ˈⁿba	=piˈa	=t͡ɕɨβ͡pkɨ̃ˈnãj	=pɨˈʔa
nou	e=ˈjasu	=ˈⁿdɨwɨ	=paˈgop	=ˈtɨβ͡p	=pɨaˈu
rădăcină	e=ˈɾao	=jaˈɻe	kʉp=kʉjoˈpɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕaˈt͡iə	=ɾaˈpɔ
piele	ˈe=ɨ	=ˈkʰʌ	=ˈpɛ	=ˈt͡ɕaj	=ˈpɨt
coadă	e=ˈɾoko͡i	=ˈⁿbɨ	=d͡ʒ=oˈkʷaj	=nãː=ˈkɨβ͡p	=ˈɾaj
alb	ˈe=ʃɨ̃	=jaˈkʰa	=d͡ʒiˈɾa	=β͡pˈdouɥ	=ˈsĩŋ
aripă	e=ˈheo	=jaˈɻa	=pɛˈo	=ɲĩˈmãɯ	=pɛˈpɔ, =jɨˈwa

Mai jos se dau doi arbori construiți pe baza aceloraşi liste. Unele date (numele limbilor și distanțele lexicostatistice) lipsesc. Completați spațiile libere. Pentru fiecare arbore, specificați dacă este manual sau automat, precum și dacă a fost generat utilizând algoritmul A sau B.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ Adnotările manuale și indicii de stabilitate au fost omişi în mod intenționat în această cerință.

(M) (10 puncte) Procedurile automate bazate pe clasele Dolgopolsky pot produce rezultate incorecte. În acest exemplu, procedura automată detectează mai multe asemănări între siriono și o anumită limbă macro-ge (khĩsêtjê) decât între siriono și alte limbi tupí. Propuneți și descrieți *pe scurt* o procedură automată modificată care ar produce o clasificare corectă a limbilor macro-ge și tupí pornind de la listele de mai sus.

⚠ Această cerință va fi evaluată doar în eventualitatea unei egalități între cele mai bune echipe.

Autorii le mulțumesc Alejandrei Vidal, Mariei Konoshenko, lui Ilya Gruntov și lui Jamthô Suyá pentru răspunsurile la întrebări despre anumite limbi. *—Andrei Niculin, Milena Veneva*

Redactori: Ivan Derjanski (redactor tehnic), Hugh Dobbs, Stanislav Gurevici, Boris Iomdin, Liam McKnight, Andrei Niculin (redactor-șef), Aleksejs Pegușevs, Jan Petr, Alexandr Piperski, Maria Rubinștein, Milena Veneva, Elysia Warner.

Textul în română: Dan-Mircea Mirea.

Succes!