Dvacátá první Mezinárodní olympiáda v lingvistice

Brasília (Brazílie), 23.–31. července 2024

Úloha soutěže družstev

Lexikostatistika je skupina metod navrhnutá pro odhadování toho, jak blízce jsou si jakékoliv jazyky navzájem příbuzné na základě jejich slovní zásoby. Tyto metody jsou obvykle použity na dlouhé seznamy slov ručně okomentovaných experty, kteří označí, jestli konkrétní pár slov domnívaně pochází ze stejného zdroje. Občas však lingvisté použijí lexikostatistické metody na seznamy slov okomentované pomocí automatizovaných procedur. Jedna taková procedura je založená na konceptu *souhláskových tříd*, zavedených sovětsko-izraelským lingvistou Aharonem Dolgopolským v roce 1964.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ɰ	Y.	j ç (na počátku kořene)	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 (na počátku kořene)	N.	n ɳ ɲ ŋ
S.	s z ʃ ʒ ʂ ʐ ɕ ʑ c ɟ					Q.	t͡ɬ d͡ɮ
H.	ħ ʕ ʜ ʢ ʰ h ɦ ʔ, samohlásky a j ç w ʍ (s výjimkou počátku kořene)						

Dolgopolského souhláskové třídy

Níže naleznete okomentované fragmenty seznamů slov z několika světových jazykových rodin. Komentáře jsou dány jako číslice v dolním indexu. Na základě těchto seznamů byly sestrojeny rodinné stromy jazyků pomocí dvou zjednodušených verzí takzvaného *StarlingNJ* algoritmu, a ke každému slovu byl přiřazen *koeficient stability*. Stromy a koeficienty na vrchu jsou založeny na ručně okomentovaných seznamech slov, a ty na spodu jsou založeny na seznamech, které byly okomentovány automaticky. Pro každý seznam slov jsou sestrojeny dva stromy: pro algoritmus A a algoritmus B. Povšimněte si, že v některých případech je možné sestrojit více různých stromů; v takových případech byl náhodně zvolen jen jeden z nich. Ke každému uzlu na každém stromu je přiřazena lexikostatistická vzdálenost. Čím větší vzdálenost, tím blížší je příbuzenství mezi jazyky. Přesnější než „lexikostatistická vzdálenost“ by tudíž byl pojem „převrácená lexikostatistická vzdálenost“. Pro jednoduchost budeme v této úloze používat pojem „lexikostatistická vzdálenost“.

Jak koeficienty stability, tak lexikostatické vzdálenosti jsou zaokrouhleny na dvě desetinná místa. Pokud je třetí číslice za desetinnou čárkou menší než 5, zaokrouhlujeme dolů; v opačném případě zaokrouhlujeme nahoru. Například, 2,836 je zaokrouhleno na 2,84, 0,705 je zaokrouhleno na 0,71, a 0,703 je zaokrouhleno na 0,70. Zaokrouhlení se týká pouze hodnot ukázaných lidským čtenářům. Jinak řečeno: počítač, který vykonává algoritmus, „vidí“ nezaokrouhlené hodnoty.

Povšimněte si, že o některých slovech se ví, anebo je předpokládáno, že byla přejmuta z jiných jazyků. Například, slovo **jokːi** ‚sůl‘ jazyka kadiwéu je přejímka od slova **jukɨ** v jazyce guaraní, a **ʔaːnʲ** ‚rok‘ jazyka ’iipay (Mesa Grande) je přejmuto ze španělského **ˈaɲo**.

V některých případech je k jednomu významu uvedeno několik synonym oddělených čárkou. Příkladem je ‚chodidlo‘ v jazyce vejoz.

V datech níže jsou předpony odděleny znakem „=“, a všechny přípony jsou odděleny znakem „-“. Některá slova se používají výhradně s předponami. Ta začínají znakem „=“.

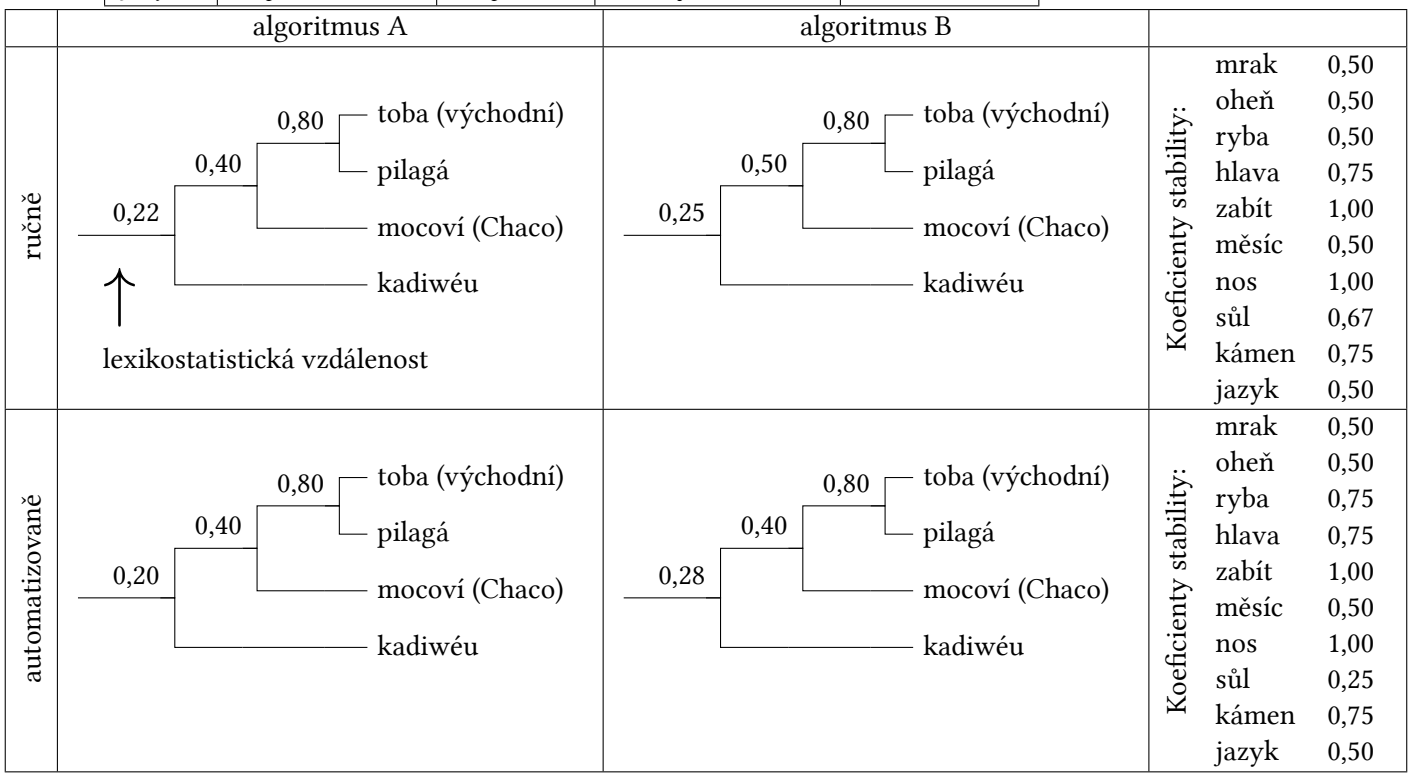
Data jsou přepsána pomocí mezinárodní fonetické abecedy. ˈ = hlavní přízvuk, ˌ = vedlejší přízvuk (slabší než hlavní přízvuk), ◌ː = dlouhá hláska, ◌̆ = velmi krátká hláska, $\widehat{XY}$ = X a Y jsou vysloveny jako jedna hláska, ◌́ = vysoký tón, ◌̀ = nízký tón, ◌̂ = klesavý tón, ◌ˀ = předglotalizovaná hláska (předcházená krátkým zahrazením proudu vzduchu v hrdle), ◌’ = ejektivní souhláska (vyslovovaná krátkým zahrazením proudu vzduchu v hrdle), ◌̥ = neznělá hláska, ◌̃ = nasalizovaná hláska (vyslovovaná skrze nos),

◌̰ = třepená fonace (hluboký, skřípavý zvuk), ⁿ◌ označuje proudění vzduchu skrze nos před souhláskou, ◌ʰ = souhláska s přídechem (vyslovená s výdechem vzduchu), ◌ʷ = labializovaná souhláska (vyslovená se zaokrouhlenými rty), ◌ʲ = palatalizovaná hláska (vyslovována s jazykem posunutým blíže tvrdému patru). **ɑ**, **æ**, **ɛ**, **ɪ**, **ɨ**, **ɔ**, **ʊ**, **ʉ**, **ə**, **ʌ**, **ɒ**, **ɘ**, **ɣ**, **ɵ**, **ø** jsou samohlásky. Zbylé speciální znaky jsou souhlásky.

⚠ Znalost jakéhokoliv z jazyků zmíněných v této úloze nepřináší žádnou výhodu při jejím řešení.

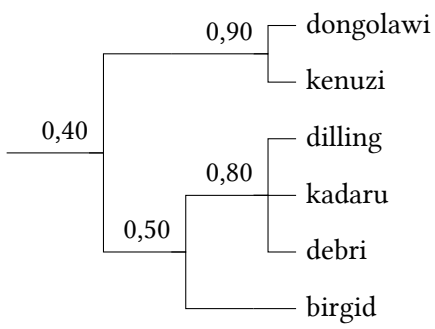
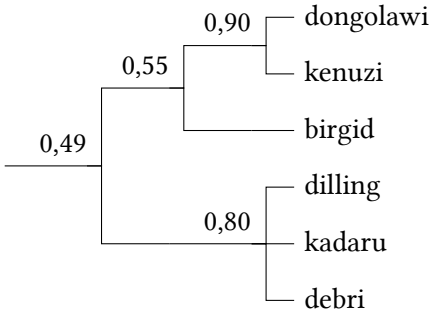
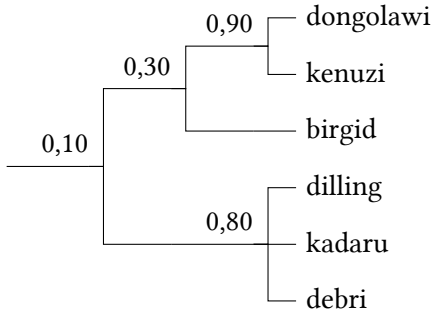
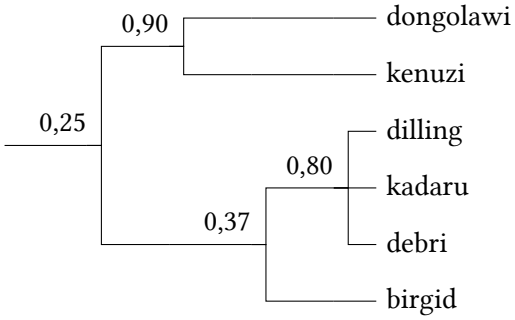
Část I. Rodina gvajkurú (Argentina, Brazílie, Paraguay)

	toba (východní)	pilagá	mocoví (Chaco)	kadiwéu
mrak	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lolːadi$_3$
oheň	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=olːedi$_2$
ryba	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nijːoɢo-d͡ʒegi$_3$
hlava	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=akːilo$_2$
zabít	=alawat$_1$	=aˈlaːt$_1$	=alawat$_1$	=elːowadi$_1$
měsíc	ʔawoʁojk$_1$	ʔaˈwoʃojk$_1$	ʃiɾajɣo$_2$	epːenaj$_3$
nos	=mik$_1$	=ˈmik$_1$	=mik$_1$	=mːiqːo$_1$
sůl	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jokːi$_{-1}$
kámen	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wetːiɢa$_2$
jazyk	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʃat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=okːelːi$_3$



Část II. Núbijská rodina (Egypt, Súdán)

	dongolawi	kenuzi	dilling	kadaru	debri	birgid
zabít	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	fila:l-e$_1$
měsíc	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	ma:l$_2$
voda	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
dát	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
dobrý	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
vítr	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
vlasy	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
břicho	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
spát	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
slunce	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	algoritmus A	algoritmus B	
ručně			Koeficienty stability: zabít 0,50 měsíc 0,83 voda 1,00 dát 1,00 dobrý 0,50 vítr 0,50 vlasy 0,83 břicho 0,83 spát 0,83 slunce 0,50
automatizovaně			Koeficienty stability: zabít 0,33 měsíc 0,50 voda 0,50 dát 0,67 dobrý 0,50 vítr 0,50 vlasy 0,83 břicho 1,00 spát 0,50 slunce 0,50

(A) **(2 body)** Souhláska **ʁ** je vyslovována jako francouzské *r*, na zadní části jazyka. Do které Dolgopolského třídy náleží, a jak jste na to přišli?

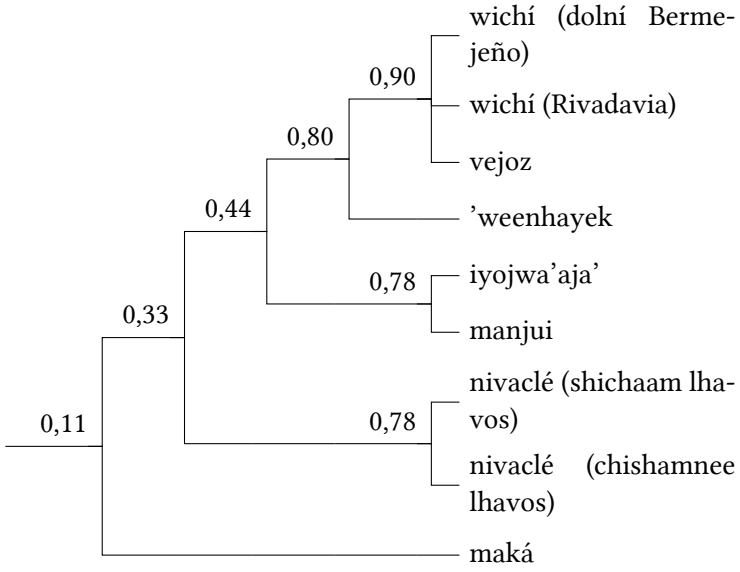
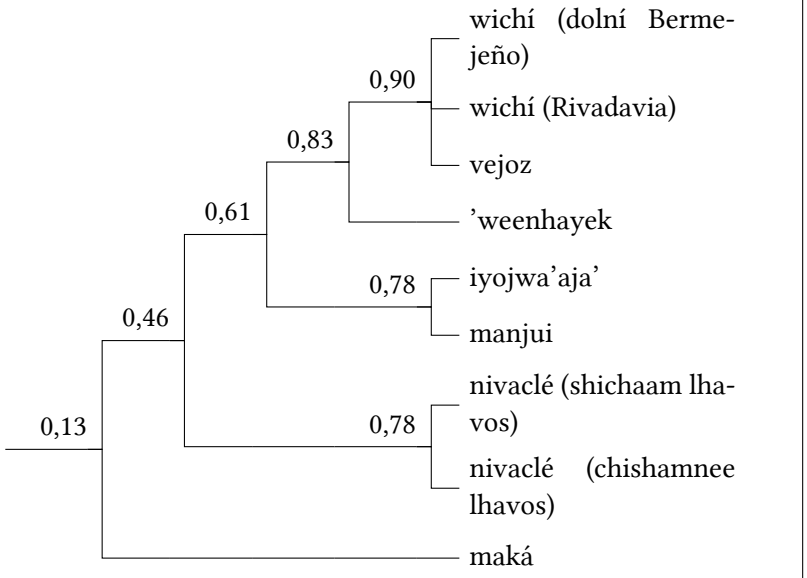
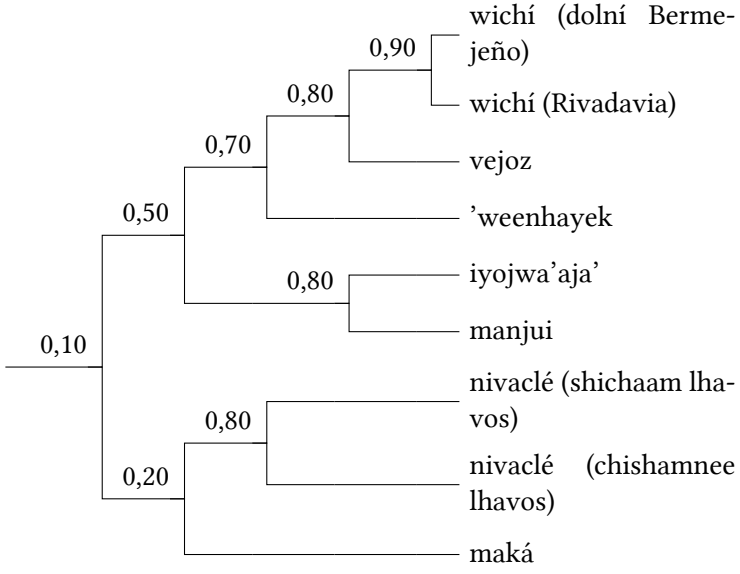
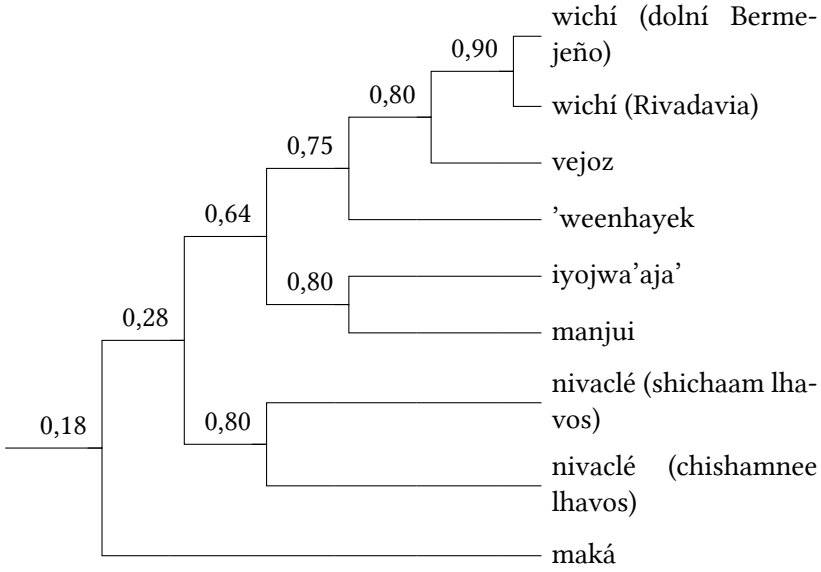
(B) **(2 body)** Nubijský strom vlevo nahoře je pouze jedním ze dvou možných stromů pro tuto kombinaci algoritmu a typu komentáře. Nakreslete druhý možný strom.

(C) **(2 body)** Nubijský strom vlevo dole je pouze jedním ze dvou možných stromů pro tuto kombinaci algoritmu a typu komentáře. Nakreslete druhý možný strom.

(D) **(2 body)** Lexikostatistická vzdálenost 0,49 (přiřazená kořenu núbijskému stromu vpravo nahoře) byla zaokrouhlena na dvě desetinná místa, stejně jako ostatní vzdálenosti v této úloze. Jaká je přesná vzdálenost?

Část III. Matagvajská rodina (Argentina, Bolívie, Paraguay)

	wichí (dolní Bermejeño)	wichí (Ri-vadavia)	vejoz	’weenhayek	iyojwa’aja’	manjui	nivaclé (shichaam lhavos)	nivaclé (chisham-nee lhavos)	maká
oheň	ʔitoχ$_{1}$	ʔitɔx$_{1}$	ʔitɑh$_{1}$	ʔiːtɑχ$_{1}$	ˈh͡wat$_{2}$	ˈʔe͡itʲe$_{1}$	ʔitɑx$_{1}$	ʔitɑx$_{1}$	feʔt$_{2}$
ryba	ʔwahat$_{1}$	wahat$_{1}$	wahat$_{1}$	ʔwaːhat$_{1}$	siˈʔjus$_{-1}$	ʃiˈʔjus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	seheț͡s$_{-1}$
chodidlo	=pat͡ʃu$_{1}$	=qɔlɔ$_{2}$	=pat͡ʃo$_{1}$, =kɑla$_{2}$	=paːkʲ’oʔ$_{1}$	=ˈsat$_{3}$	=kaˈʔlaʔ$_{2}$	=ɸoʔ$_{4}$	=ɸoʔ$_{4}$	=fˈ’iʔ$_{5}$
voda	ʔinot$_{1}$	ʔinɔt$_{1}$	wah$_{2}$	ʔinɑːt$_{1}$	ʔiˈnʲat$_{1}$	ʔaˈʔnat$_{1}$	jinɑʔt$_{1}$	jinɑʔt$_{1}$	iweliʔ$_{3}$
dát	=ʔweɲ̥-u$_{1}$	=weɲ̥-u$_{1}$	=ʔweɲ̥-o$_{1}$	=ʔweɲ̥-oʔ$_{1}$	=ˈwɛh͡n-aʔm$_{2}$	=ˈhajʔ$_{3}$, =ˈwɛn$_{2}$	=xut$_{4}$	=xut-ej$_{4}$	tis-ix$_{5}$
dobrý	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ˈʔes$_{1}$	ˈʔe͡is$_{1}$	ʔis$_{1}$	ʔis$_{1}$	t=ejk’un-ej$_{2}$
vítr	ʔinwokʷ$_{1}$	ʔinwɔk$_{1}$	ʔihwokʷ$_{1}$	=jaːɬ$_{2}$, =xʷoxʷ$_{3}$	ˈh͡lah͡wuʔ$_{4}$	ˈh͡lah͡wuuʔ$_{4}$	ɬaβiʔm$_{5}$	ɬaβiʔm$_{5}$	t’unik’i$_{6}$
strom	haʔlo$_{1}$	halɔ$_{1}$	haʔlɑ$_{1}$	haʔlɑʔ$_{1}$	ʔaˈʔlaʔ$_{1}$	ʔaˈʔla-k$_{1}$	ʔaʔkxi-juk$_{2}$	jiʔk͡lɑʔ$_{1}$	naxka-k$_{3}$
vlasy	=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$_{5}$
zabít	=lon$_{1}$	=lɔn$_{1}$	=lɑn$_{1}$	=lɑːn̥$_{1}$	=ˈlaʔan$_{1}$	=ˈlan$_{1}$	=k͡lɑn$_{1}$	=k͡lɑn$_{1}$	=lan$_{1}$

	algoritmus A	algoritmus B	
ručně			Koeficienty stability: oheň 0,78 ryba 1,00 chodidlo 0,33 voda 0,78 dát 0,44 dobrý 0,89 vítr 0,33 strom 0,78 vlasy 0,67 zabít 1,00
automatizovaně			Koeficienty stability: oheň 0,78 ryba 0,44 chodidlo 0,33 voda 0,56 dát 0,67 dobrý 0,89 vítr 0,22 strom 0,67 vlasy 0,67 zabít 1,00

Část IV. Mongolská rodina (Čínská lidová republika, Mongolsko, Rusko)

(E) (10 bodů) Prozkoumejte následující seznam slov. Spočítejte koeficienty stability odpovídající jak ručním, tak automatizovaným komentářům.

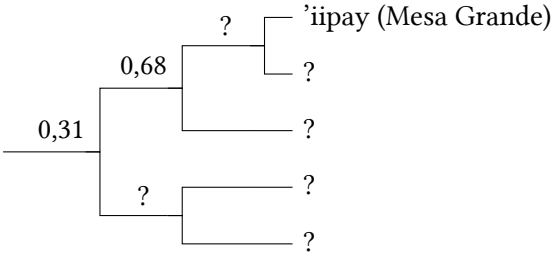
Abychom Vám pomohli, spočítali jsme oba koeficienty stability pro slovo ‚vše'. V náhodnám pořadí se jedná o 0,36 a 0,40.

	dagurština (hailarský dialekt)	chamnigan-ština (man-džuský dialekt)	burjatština (chorijský dialekt)	novo-bargut-ština	öeletština	chošúd-ština	kalmyčtina	chalcha-ština	ordoština	šira-ju-gurština	bonanština
vše	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$_{-2}$
kůra	hai̯l͡s$_{1}$	qalihʊn$_{1}$	χəltəhən$_{2}$	xalʲhuː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xəɮt^{h}ɔ̆s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
břicho	keːli$_{1}$	gʊtəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
pták	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡zer$_{2}$
oheň	ɢali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
cesta	terg-uːl$_{1}$	qarɢʊi̯$_{2}$	χarɢɨ$_{2}$, zam$_{-1}$	zam$_{-1}$	d͡zam$_{-1}$	d͡zam$_{-1}$	xaː-lʁə$_{3}$	t͡sam$_{-1}$	t͡ʃam$_{-1}$	mør$_{4}$	mor$_{4}$
sůl	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
plavat	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-$_{1}$
voda	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
vítr	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$_{1}$

Část V. Jumanská rodina (Mexiko, Spojené státy americké)

(F) (8 bodů) Prozkoumejte následující seznam slov. Níže je k vidění strom, který byl sestaven na základě stejného seznamu. Některá data (názvy jazyků a lexikostatické vzdálenosti) chybí. Vyplňte mezery. Stanovte, jestli byl strom sestaven ručně, nebo automaticky, a jestli byl vytvořen pomocí algoritmu A, nebo B.

	mojave	cocopa	yavapai	tiipay (jamul)	ʼiipay (Mesa Grande)
krátký	wena=wen-a_1	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i_2	ləˈʔuɲ$_1$	mə=put-k_3
pták	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔ=ˈʃa$_2$	ʔaː=ʂaː$_2$
kost	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a_1	ˈak$_1$	aq$_1$
suchý	i=roː-v-k_1	ˈʂ=ʔarˌ$_2$	ˈru-β-i_1	ˈs=ʔaːj$_3$	saːj$_3$
maso	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeːˈθo-β-a_3	ˈkʷak$_4$	kukʷaːj-p_1
krk	maʌaqe$_1$	ˈm=puk$_2$	ˈmlqí$_1$	iːˈpuk$_2$	iː=puk$_2$
vidět	i=juː-k_1	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
ocas	iː=ʔar$_1$	ˈʃ=juʌ̊$_2$	ˈβ=hé$_3$	ʃəˈjuɬ$_2$	xə=juɬ$_2$
dvě	havik-k_1	ˈx=wak$_1$	ˈh͡wâk-i_1	xəˈwak$_1$	xə=wak$_1$
rok	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a_3	matˈwam$_2$	ʔaːnʲ$_{-1}$



(G) (20 bodů) Několik dalších stromů bylo vytvořeno z jumanských jazyků, s následujícími lexikostatistickými vzdálenostmi u kořene stromu (tj. lexikostatickými vzdálenostmi na levém kraji stromu):

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

Nakreslete každý z těchto stromů. Pro každý ze stromů stanovte, jestli byl sestaven ručně, nebo automaticky, a jestli byl vytvořen pomocí algoritmu A, nebo B.

(H) (3 body) Dvě ze vzdáleností uvedených v Úkolu **(G)** byly zaokrouhleny na dvě desetinná místa: 0,23 bylo zaokrouhleno z 0,225. Jaká další vzdálenost byla zaokrouhlena, a jaká je její přesná hodnota?

(I) (4 body) Vysvětlete, jak jsou počítány koeficienty stability.

(J) (5 bodů) Vysvětlete, jak jsou počítány lexikostatistické vzdálenosti.

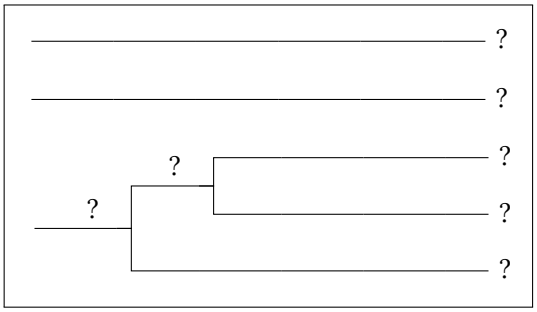
(K) (4 body) Vysvětlete rozdíl mezi algoritmem A a algoritmem B.

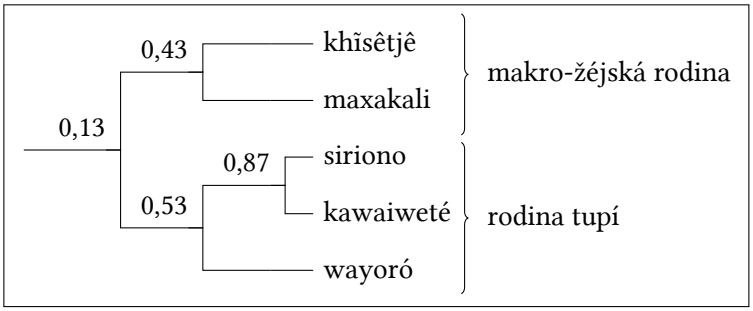
Část VI. Makro-žéjská rodina a rodina tupí (Brazílie, Bolívie)

(L) **(28 bodů)** Makro-žéjská rodina a rodina tupí jsou dvě z hlavních jazykových rodin v Jižní Americe. Někteří lingvisté věří, že jsou vzdáleně spřízněné. Prozkoumejte následující seznamy slov.

	A	B	Γ	Δ	E
kůra	e=ˈe-ke	hʷĩ=ˈkʰʌ	kʉp=ˈpɛ	mĩβ͡m=ˈt͡ɕɑj	=ˈpɛ
břicho	ˈe=ɾje	=ˈtʰigi	=ãˈʉ̃n	=ˈtæj	=ɾɛˈwɛk
krev	e=ˈɾuki	=kaˈⁿbɽo	=d͡ʒ=aˈʉ	=ˈhɛβ͡p	=ɾuˈɨ
pálit	=ˈɾai͡	=ɾɔ=ˈkʰɹə̃	=poˈkʷa	mũ=...=ˈhaβ͡p	=ɾaˈpɨ
tuk	e=ˈkiɾa	=ˈtʰwɜmɨ	=ˈd͡ʒ=ap	=ˈtuβ͡p	=ˈkap
chodidlo	ˈe=i	=ˈhʷaji	=ˈβi	=pɒˈta	=ˈpɨ
ruka	ˈe=o	=ɲɨ̃ˈkʰɹa	=ˈβo	=ˈɲĩβ͡m	=ˈpɔ
těžký	e=ˈusi	=wɨˈtʰĩ	=poˈti	=β͡pˈtəj	=pɔˈɨj
játra	ˈe=ja	=ˈⁿba	=piˈa	=t͡ɕɨβ͡pkɨ̃ˈnãj	=pɨˈʔa
nový	e=ˈjasu	=ˈⁿdɨwɨ	=paˈgop	=ˈtɨβ͡p	=pɨaˈu
kořen	e=ˈɾao	=jaˈtɾe	kʉp=kʉjoˈpɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕaˈtɨ͡ə	=ɾaˈpɔ
kůže	ˈe=ɨ	=ˈkʰʌ	=ˈpɛ	=ˈt͡ɕɑj	=ˈpit
ocas	e=ˈɾoko͡i	=ˈⁿbɨ	=d͡ʒ=oˈkʷaj	=nãː=ˈkɨβ͡p	=ˈɾaj
bílý	ˈe=ʃɨ̃	=jaˈkʰa	=d͡ʒiˈɾa	=β͡pˈdoɯ	=ˈsĩŋ
křídlo	e=ˈheo	=jaˈtɾa	=pɛˈo	=ɲĩˈmãɯ	=pɛˈpɔ, =jɨˈwa

Níže jsou k vidění dva stromy, které byly sestaveny na základě stejných seznamů. Některá data (názvy jazyků a lexikostatické vzdálenosti) chybí. Vyplňte mezery. Pro každý ze stromů stanovte, jestli byl sestaven ručně, nebo automaticky, a jestli byl vytvořen pomocí algoritmu A, nebo B.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ Ruční komentáře a koeficienty stability byly záměrně vynechány v tomto úkolu.

(M) **(10 bodů)** Automatizované procedury založené na Dolgopolského třídách mohou vést k nesprávným výsledkům. V tomto příkladě automatizovaná procedura detekuje více podobností mezi jazykem siriono a určitým makro-žéjským jazykem (khĩsêtjê), než mezi jazykem siriono a ostatními tupíjskými jazyky. Navrhněte a *stručně* popište upravenou automatizovanou proceduru, která by vedla ke správné klasifikaci, pokud by byla použita na makro-žéjské a tupíjské seznamy slov výše.

⚠ Tento úkol bude ohodnocený pouze v případě remízy mezi týmy s nejvyšším počtem bodů.

Autoři děkují Alejandře Vidal, Marii Konoshenko, Ilyovi Gruntovovi a Jamthô Suyá za odpovědi na dotazy ohledně konkrétních jazyků. *—Andrej Nikulin, Milena Venevová*

Redakce: Ivan Deržanski (techn. red.), Hugh Dobbs, Stanislav Gurevič, Boris Iomdin, Liam McKnight, Andrej Nikulin (odp. red.), Aleksejs Peguševs, Jan Petr, Aleksandr Piperski, Marija Rubinštejnová, Milena Venevová, Elisija Warnerová.

Český text: Jan Petr.

Hodně štěstí!