Vingt-et-unièmes Olympiades internationales de linguistique

Brasília (Brésil), du 23 au 31 juillet 2024

Problème de la compétition en équipe

La lexicostatistique est un groupe de méthodes créées pour estimer la proximité entre différentes langues à base de leurs vocabulaires. Ces méthodes sont normalement appliquées à de longues listes de mots annotées par des experts, qui indiquent si l'on pense qu'une paire de mots spécifique provient de la même source. Cependant, les linguistes appliquent parfois des méthodes lexicostatistiques aux listes de mots annotées au moyen de processus automatisés. Un tel processus est basé sur le concept des *classes de consonnes*, introduit en 1964 par le linguiste soviéto-israélien Aharon Dolgopolsky.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ʯ	Y.	j ç (au début de la racine)	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 (au début de la racine)	N.	n ɳ ɲ ŋ
S.	s z ʃ ʒ ʂ ʐ ɕ ʑ ɕ ʝ					Q.	t͡ɬ d͡ɮ
H.	ħ ʕ ʜ ʢ ʔ h ɦ ʔ, les voyelles, et j ç w ʍ (sauf au début de la racine)						

Les classes de consonnes de Dolgopolsky

Ci-dessous se trouvent des fragments annotés de listes de mots de quelques familles de langues du monde. Les annotations sont données au moyen de chiffres en indice. Sur la base de ces listes, des arbres généalogiques des langues ont été construits en utilisant deux versions simplifiées de l'algorithme appelé *StarlingNJ*, et un *indice de stabilité* a été assigné à chaque mot. Les arbres et indices de stabilité en haut sont basés sur des listes de mots annotées manuellement, et ceux en bas sont basés sur des listes qui ont été annotées automatiquement. Deux arbres ont été construits pour chaque liste de mots, en utilisant deux versions de l'algorithme : l'Algorithme A et l'Algorithme B. Faites attention que dans certains cas il y a plusieurs arbres possibles qui correspondent à une seule liste de mots; dans ces cas un seul arbre a été choisi aléatoirement. Une distance lexicostatistique a été assignée à chaque nœud de chaque arbre. Plus la distance est grande, plus la proximité entre les langues est immédiate. Un terme plus précis est donc «distance lexicostatistique inverse» plutôt que «distance lexicostatistique». Pour simplifier, le terme «distance lexicostatistique» est utilisé dans ce problème.

Les indices de stabilité ainsi que les distances lexicostatistiques ont été arrondis à deux décimales. Si le troisième chiffre après la virgule est inférieur à 5, arrondissez à la valeur inférieure; sinon, arrondissez à la valeur supérieure. Par exemple, 2,836 est arrondi à 2,84, 0,705 à 0,71, et 0,703 à 0,70. L'arrondi ne s'applique qu'aux valeurs présentées aux lecteurs humains. Autrement dit l'ordinateur qui exécute les algorithmes peut «voir» les valeurs non-arrondies.

Notez que l'on sait ou estime que quelques mots ont été empruntés aux autres langues. Par exemple, le mot **jokːi** 'sel' de la langue kadiwéu est emprunté au mot guarani **jukɨ**, et **ʔaːnʲ** 'année' du ipai (de Mesa Grande) est emprunté au mot espagnol **ˈaɲo**.

Dans certains cas plusieurs synonymes d'un seul sens sont donnés dans les listes de mots, séparés par des virgules. Un exemple est 'pied' de la langue vejoz.

Dans les données ci-dessous, tous les préfixes sont séparés par le signe «=» et tous les suffixes sont séparés par le signe «-». Certains mots ne se trouvent qu'avec des préfixes. Ces mots commencent par le signe «=».

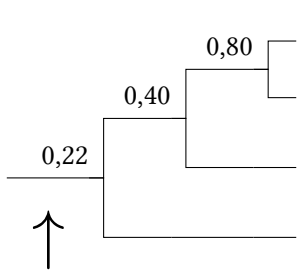
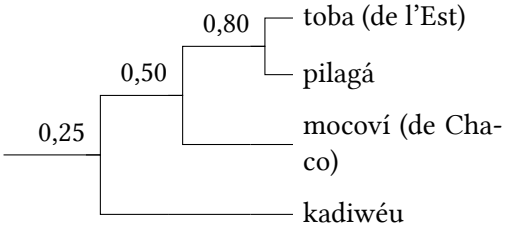
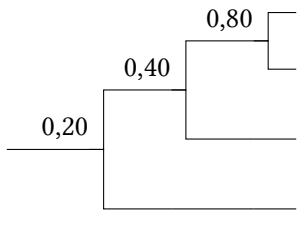
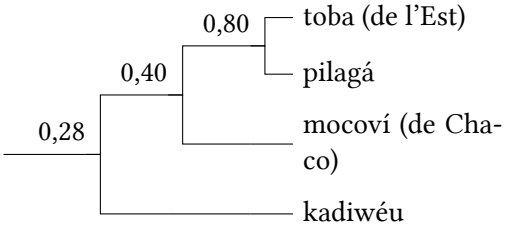
Les données ont été transcrites dans l'Alphabet Phonétique International. ˈ = accent primaire, ˌ = accent secondaire (plus faible que l'accent primaire), ○ː = son long, ○̆ = son très court, X͡Y = X et Y sont prononcés comme un seul son, ○́ = ton haut, ○̀ = ton bas, ○̂ = ton descendant, ʔ○ = son préglottalisé

(précédé par un bref blocage du passage d'air dans la gorge), ◌ʼ = son éjectif (prononcé en bloquant brièvement le passage d'air dans la gorge), ◌̥ = son sourd, ◌̃ = son nasalisé (prononcé avec passage de l'air par le nez), ◌̰ = voix craquée (tonalité basse et rauque), ⁿ◌ indique le flux d'air par le nez avant la consonne, ◌ʰ = consonne aspirée (prononcée avec une bouffée d'air), ◌ʷ = consonne labialisée (prononcée avec lèvres arrondies), ◌ʲ = son palatalisé (prononcé en dirigeant une partie de la langue près du palais dur). **ɑ**, **æ**, **ɛ**, **ɪ**, **ɨ**, **ɔ**, **ʊ**, **ʉ**, **ə**, **ʌ**, **ɒ**, **ɘ**, **ʏ**, **ɵ**, **ø** sont des voyelles. Les autres caractères spéciaux sont des consonnes.

⚠ Toute connaissance des langues mentionnées dans le problème ne donne aucun avantage pour résoudre le problème.

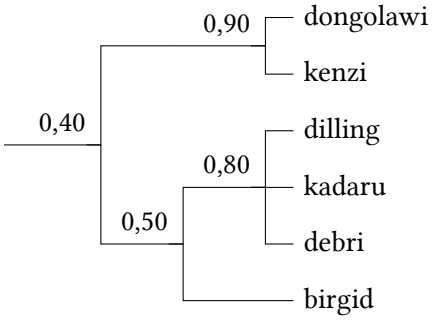
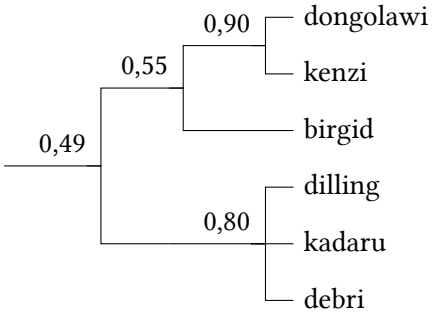
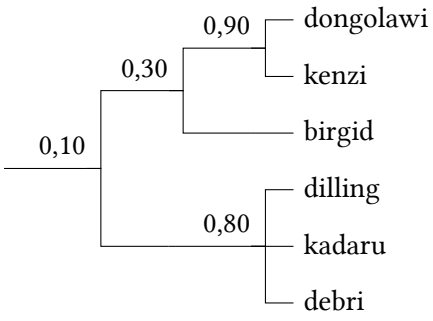
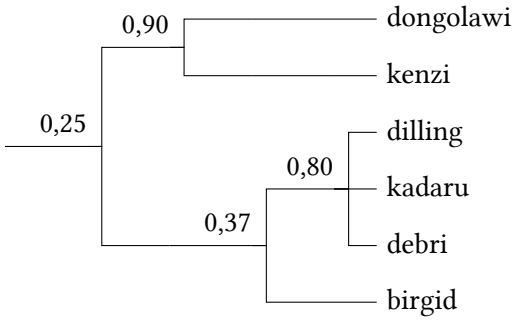
I[ère] partie. La famille waykuruane (Argentine, Brésil, Paraguay)

	toba (de l'Est)	pilagá	mocoví (de Chaco)	kadiwéu
nuage	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lolːadi$_3$
feu	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=olːedi$_2$
poisson	njaq$_1$	ˈnijaq$_1$	naʃin$_2$	nijːoɢo-d͡ʒegi$_3$
tête	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=akːilo$_2$
tuer	=alawat$_1$	=aˈlaːt$_1$	=alawat$_1$	=elːowadi$_1$
lune	ʔawoʁojk$_1$	ʔaˈwoʃojk$_1$	ʃɪrajɣo$_2$	epːenaj$_3$
nez	=mik$_1$	=ˈmik$_1$	=mik$_1$	=mːiqːo$_1$
sel	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jokːi$_{-1}$
pierre	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wetːiɢa$_2$
langue	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʃat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=okːelːi$_3$

	algorithme A	algorithme B	
manuel			Indices de stabilité : nuage 0,50 feu 0,50 poisson 0,50 tête 0,75 tuer 1,00 lune 0,50 nez 1,00 sel 0,67 pierre 0,75 langue 0,50
automatique			Indices de stabilité : nuage 0,50 feu 0,50 poisson 0,75 tête 0,75 tuer 1,00 lune 0,50 nez 1,00 sel 0,25 pierre 0,75 langue 0,50

II^{ème} partie. La famille nubienne (Egypte, Soudan)

	dongolawi	kenzi	dilling	kadaru	debri	birgid
tuer	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	filaːl-e$_1$
lune	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	maːl$_2$
eau	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
donner	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
bon	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
vent	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
cheveux	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
ventre	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
dormir	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
soleil	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	algorithme A	algorithme B	
manuel			Indices de stabilité : tuer 0,50 lune 0,83 eau 1,00 donner 1,00 bon 0,50 vent 0,50 cheveux 0,83 ventre 0,83 dormir 0,83 soleil 0,50
automatique			Indices de stabilité : tuer 0,33 lune 0,50 eau 0,50 donner 0,67 bon 0,50 vent 0,50 cheveux 0,83 ventre 1,00 dormir 0,50 soleil 0,50

(A) **(2 points)** La consonne **ʁ** est prononcée comme l'*r* grasseyé du français, c'est-à-dire à l'arrière de la langue. À quelle classe Dolgopolsky appartient-elle, et comment l'avez-vous déterminé ?

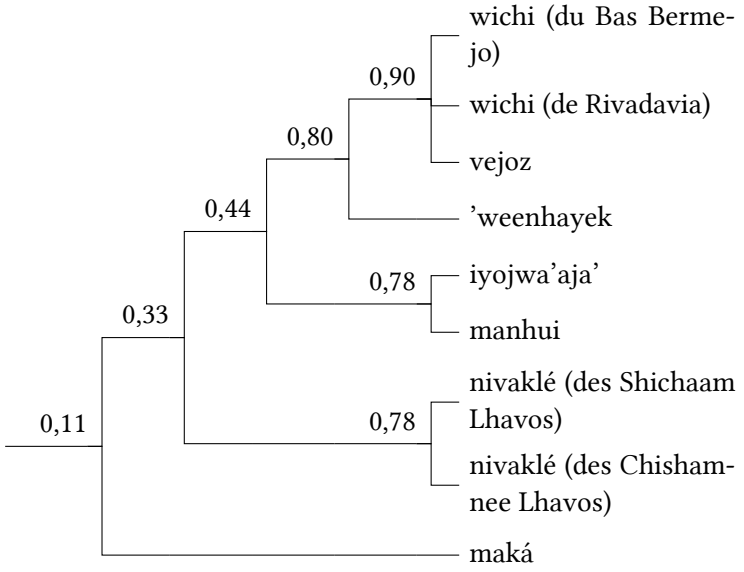
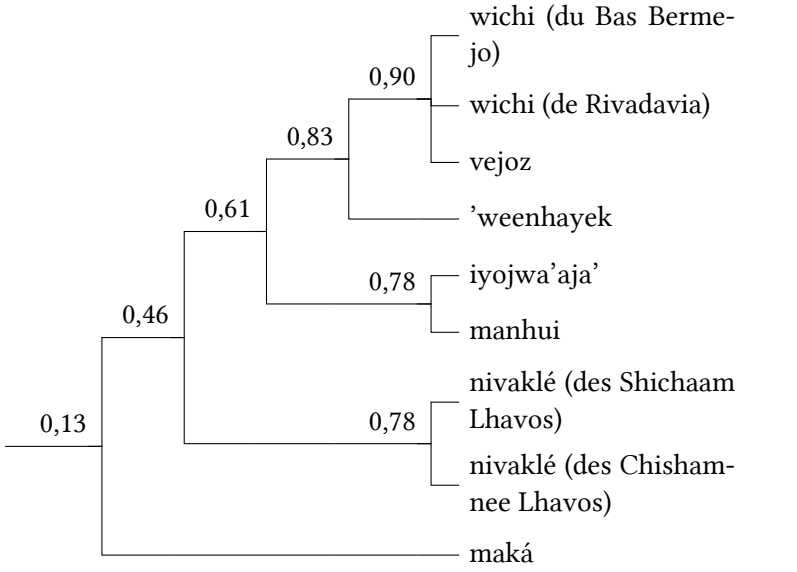
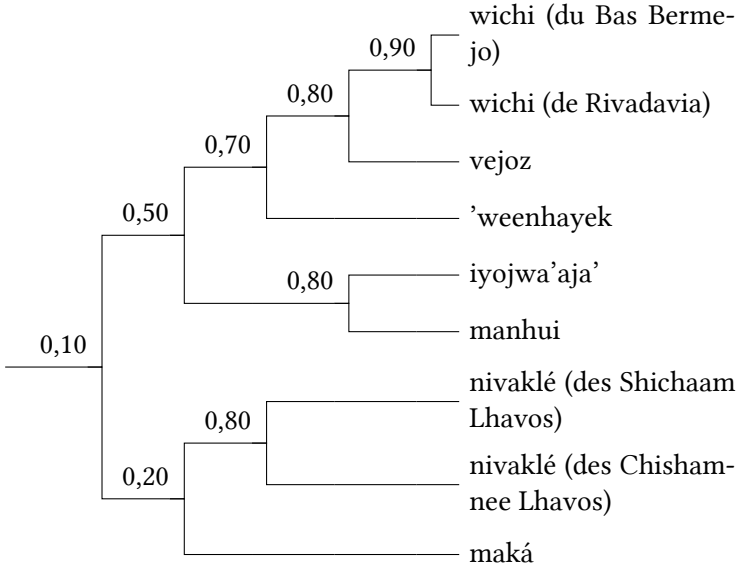
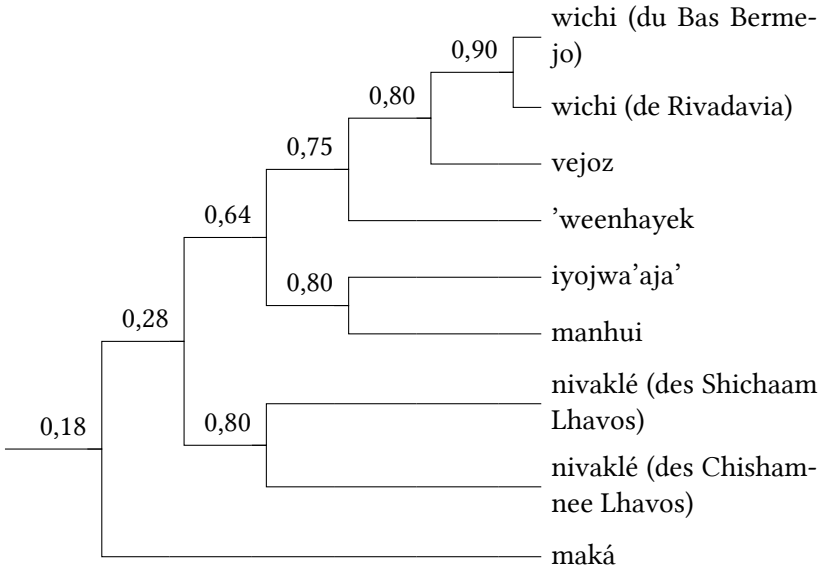
(B) **(2 points)** L'arbre nubien en haut à gauche est juste un des deux arbres possibles pour cette combinaison d'algorithme et de type d'annotation. Dessinez l'autre arbre possible.

(C) **(2 points)** L'arbre nubien en bas à gauche est juste un des deux arbres possibles pour cette combinaison d'algorithme et de type d'annotation. Dessinez l'autre arbre possible.

(D) **(2 points)** Comme quelques autres distances dans ce problème, la distance lexicostatistique 0,49 (assignée à la racine de l'arbre nubien en haut à droite) a été arrondie à deux décimales. Quelle est la distance exacte ?

III^{ème} partie. La famille mataguayo (Argentine, Bolivie, Paraguay)

	wichi (du Bas Berme-jo)	wichi (de Rivada-via)	vejoz	’weenhayek	iyojwa’aja’	manhui	nivaklé (des Shi-chaam Lhavos)	nivaklé (des Chi-shamnee Lhavos)	maká
feu	ʔitoχ$_1$	ʔitɔx$_1$	ʔitɑh$_1$	ʔi:tɑχ$_1$	ˈh͡wat$_2$	ˈʔe͡itʲe$_1$	ʔitɑx$_1$	ʔitɑx$_1$	fe^{ʔ}t$_2$
poisson	^{ʔ}wahat$_1$	wahat$_1$	wahat$_1$	^{ʔ}wa:hat$_1$	siˈ^{ʔ}jus$_{-1}$	ʃiˈ^{ʔ}jus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	sehet͡s$_{-1}$
pied	=pat͡ʃu$_1$	=qɔlɔ$_2$	=pat͡ʃo$_1$, =kɑla$_2$	=pa:kʲ’oʔ$_1$	=ˈsat$_3$	=kaˈ^{ʔ}laʔ$_2$	=ɸoʔ$_4$	=ɸoʔ$_4$	=f’iʔ$_5$
eau	ʔinot$_1$	ʔinɔt$_1$	wah$_2$	ʔinɑ:t$_1$	ʔiˈnʲat$_1$	ʔaˈ^{ʔ}nat$_1$	jinɑʔt$_1$	jinɑʔt$_1$	iweliʔ$_3$
donner	=^{ʔ}wen̪-u$_1$	=wen̪-u$_1$	=^{ʔ}wen̪-o$_1$	=^{ʔ}wen̪-oʔ$_1$	=ˈwɛh͡n-aʔm$_2$	=ˈhajʔ$_3$, =ˈwɛn$_2$	=xut$_4$	=xut-ej$_4$	tis-ix$_5$
bon	ʔis$_1$	ʔis$_1$	ʔis$_1$	ʔis$_1$	ˈʔes$_1$	ˈʔe͡is$_1$	ʔis$_1$	ʔis$_1$	t=ejk’un-ej$_2$
vent	ʔinwokʷ$_1$	ʔinwɔk$_1$	ʔihwokʷ$_1$	=ja:ɬ$_2$, =xʷoxʷ$_3$	ˈh͡lah͡wuʔ$_4$	ˈh͡lah͡wuuʔ$_4$	ɬaβiʔm$_5$	ɬaβiʔm$_5$	t’unik’i$_6$
arbre	haʔlo$_1$	halɔ$_1$	haʔlɑ$_1$	haʔlɑʔ$_1$	ʔaˈ^{ʔ}laʔ$_1$	ʔaˈ^{ʔ}la-k$_1$	ʔaʔkxi-juk$_2$	jiʔk͡lɑʔ$_1$	naxka-k$_3$
cheveux	=^{ʔ}wule-j$_1$	=wule-j$_1$	=^{ʔ}wole-j$_1$	=^{ʔ}wo:le-ç$_1$, hi:lenaχ$_2$	=ˈ^{ʔ}wole$_1$	=ˈ^{ʔ}wole-j$_1$	=ʃateʔt͡ʃ$_3$	=jeʔs$_4$	=ʔewkux-it͡s$_5$
tuer	=lon$_1$	=lɔn$_1$	=lɑn$_1$	=lɑ:n̪$_1$	=ˈlaʔan$_1$	=ˈlan$_1$	=k͡lɑn$_1$	=k͡lɑn$_1$	=lan$_1$

<table>
<tr><th></th><th>algorithme A</th><th>algorithme B</th><th></th></tr>
<tr>
<td>manuel</td>
<td> 
</td>
<td> 
</td>
<td>Indices de stabilité :
feu 0,78
poisson 1,00
pied 0,33
eau 0,78
donner 0,44
bon 0,89
vent 0,33
arbre 0,78
cheveux 0,67
tuer 1,00</td>
</tr>
<tr>
<td>automatique</td>
<td> 
</td>
<td> 
</td>
<td>Indices de stabilité :
feu 0,78
poisson 0,44
pied 0,33
eau 0,56
donner 0,67
bon 0,89
vent 0,22
arbre 0,67
cheveux 0,67
tuer 1,00</td>
</tr>
</table>

IV^{ème} partie. La famille mongolique (République populaire de Chine, Mongolie, Russie)

(E) (10 points) Examinez la liste de mots suivante. Calculez les indices de stabilité qui correspondent aux annotations manuelles et automatisées.

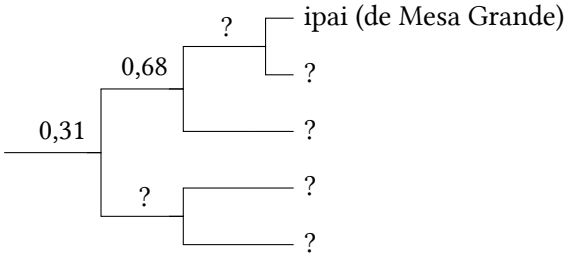
Pour vous aider, les deux indices de stabilité du mot ‘tous’ ont déjà été calculés. Dans un ordre aléatoire ils sont : 0,36 et 0,40.

	daur (de Hailar)	khamnigan (mandchou)	bouriate (de Khori)	nouveau-bargu	ööld	khochoute	kalmouk	khalkha	ordos	shira yugur	bonan
tous	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$_{-2}$
écorce	hai̯l͡s$_{1}$	qalihʊn$_{1}$	χɔltɔhɔn$_{2}$	xaɮʲhuː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xɔɮʒt^{h}ɔ̆s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
ventre	keːli$_{1}$	gətəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮʒĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
oiseau	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡ʑer$_{2}$
feu	ɢali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
chemin	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$_{4}$
sel	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
nager	unpa-du$_{1}$	ʊmba-$_{1}$	t^{h}amar-$_{2}$	umb-$_{1}$	sele-$_{3}$	umba-$_{1}$	us-t͡ɕi-$_{4}$, øːm-$_{5}$	siɮʒĭ-$_{3}$	usu-t͡ʃ^{h}i-la-$_{4}$	umpa-$_{1}$	mba-$_{1}$
eau	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
vent	kei̯n$_{1}$	halkin$_{2}$	halxin$_{2}$	halxi$_{2}$	saɮʲxin$_{2}$	salkʲi$_{2}$	saɮʲkn$_{2}$	saɮʒxĭ$_{2}$	k^{h}iː$_{1}$	k^{h}iː$_{1}$	ki$_{1}$

V^{ème} partie. La famille yumane (Mexique, États-Unis)

(F) (8 points) Examinez la liste de mots suivante. Ci-dessous est un arbre construit en utilisant la même liste. Certaines données (noms de langues et distances lexicostatistiques) manquent. Remplissez les trous. Précisez si l'arbre est manuel ou automatisé, ainsi que s'il a été généré en utilisant l'Algorithme A ou B.

	mojave	cocopa	yavapai	tipai (de Jamul)	ipai (de Mesa Grande)
court	wena=wen-a$_1$	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i$_2$	lə=ˈʔuɲ$_1$	mə=put-k$_3$
oiseau	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔ=ˈʃa$_2$	ʔaː=ʂaː$_2$
os	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a$_1$	ˈak$_1$	aq$_1$
sec	i=roː-v-k$_1$	ˈʂ=ʔaɾ$_2$	ˈru-β-i$_1$	ˈs=ʔaːj$_3$	saːj$_3$
viande	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeː=ˈθo-β-a$_3$	ˈkʷak$_4$	kukʷaːj-p$_1$
cou	maʌaqe$_1$	ˈm=puk$_2$	ˈmlqí$_1$	iː=ˈpuk$_2$	iː=puk$_2$
voir	i=juː-k$_1$	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
queue	iː=ʔar$_1$	ˈʃ=juʌ̊$_2$	ˈβ=hé$_3$	ʃə=ˈjuɬ$_2$	xə=juɬ$_2$
deux	havik-k$_1$	ˈx=wak$_1$	ˈh͡wâk-i$_1$	xə=ˈwak$_1$	xə=wak$_1$
année	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a$_3$	mat-ˈwam$_2$	ʔaːnʲ$_{-1}$



(G) (20 points) Quelques autres arbres ont été générés pour la famille yumane, avec les distances lexicostatistiques suivantes à la racine de l'arbre (les distances lexicostatistiques tout à gauche de chaque arbre) :

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

Dessinez chacun de ces arbres. Pour chacun des arbres, précisez s'il est manuel ou automatisé, ainsi que s'il a été généré en utilisant l'Algorithme A ou B.

(H) (3 points) Deux distances données dans la tâche **(G)** ont été arrondies à deux décimales : 0,23 a été arrondi à partir de 0,225. Quelle autre distance a été arrondie, et quelle est sa valeur exacte ?

(I) (4 points) Expliquez comment sont calculés les indices de stabilité.

(J) (5 points) Expliquez comment sont calculées les distances lexicostatistiques.

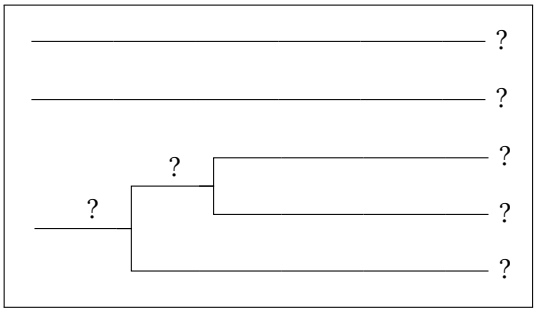
(K) (4 points) Expliquez la différence entre les algorithmes A et B.

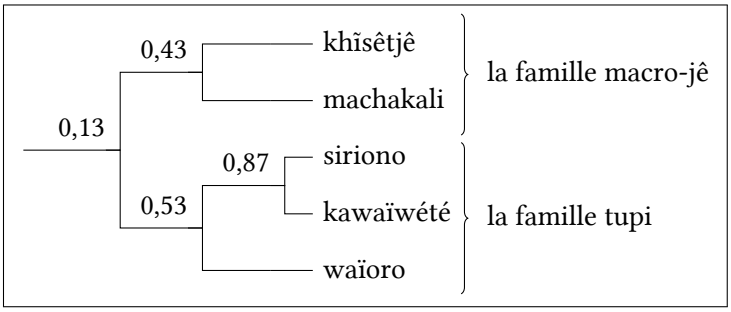
VI^{ème} partie. La famille macro-jê et la famille tupi (Brésil, Bolivie)

(L) (28 points) La famille macro-jê et la famille tupi sont deux familles de langues importantes d'Amérique du Sud. Quelques linguistes croient qu'elles sont lointainement apparentée. Examinez les listes de mots suivantes.

	A	B	Γ	Δ	E
écorce	e='e-ke	hʷĩ='kʰʌ	kʉp='pɛ	mĩβ͡m='t͡ɕaj	='pɛ
ventre	'e=ɾje	='ʈʰigi	=ã'ʉ̃n	='tæj	=ɾɛ'wɛk
sang	e='ɾuki	=ka'ⁿbɽo	=d͡ʒ=a'ʉ	='hɛβ͡p	=ɾu'ɨ
brûler	='ɾai̯͡	=ɾɔ='kʰɹə̃	=po'kʷa	mũ=...='haβ͡p	=ɾa'pɨ
graisse	e='kiɾa	='ʈʰwɜmɨ	='d͡ʒ=ap	='tuβ͡p	='kap
pied	'e=i	='hʷaji	='βi	=pɒ'ta	='pɨ
main	'e=o	=ɲĩ'kʰɹa	='βo	='ɲĩβ͡m	='pɔ
lourd	e='usi	=wɨ'ʈʰĩ	=po'ti	=β͡p'təj	=pɔ'ɨj
foie	'e=ja	='ⁿba	=pi'a	=t͡ɕɨβ͡pkĩ'nãj	=pɨ'ʔa
nouveau	e='jasu	='ⁿdɨwɨ	=pa'gop	='tɨβ͡p	=pɨa'u
racine	e='ɾao	=ja'ɽe	kʉp=kʉjo'pɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕa'tɨ͡ə	=ɾa'pɔ
peau	'e=ɨ	='kʰʌ	='pɛ	='t͡ɕaj	='pit
queue	e='ɾoko̯͡i̯	='ⁿbɨ	=d͡ʒ=o'kʷaj	=nã:='kɨβ͡p	='ɾaj
blanc	'e=ʃɨ	=ja'kʰa	=d͡ʒi'ɾa	=β͡p'douɻ	='sĩŋ
aile	e='heo	=ja'ɽa	=pɛ'o	=ɲĩ'mãuɻ	=pɛ'pɔ, =jɨ'wa

Ci-dessous sont deux arbres construits en utilisant les mêmes listes. Certaines données (noms de langues et distances lexicostatistiques) manquent. Remplissez les trous. Pour chacun des arbres, précisez s'il est manuel ou automatisé, ainsi que s'il a été généré en utilisant l'Algorithme A ou B.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ Les annotations manuelles et les indices de stabilité ont été intentionellement omis pour cette tâche.

(M) (10 points) Les processus automatisés basés sur les classes Dolgopolsky peuvent produire des résultats erronés. Dans cet exemple le processus automatisé détecte plus de similarités entre le siriono et une certaine langue macro-jê (khĩsêtjê), qu'entre le siriono et les autres langues tupi. Proposez un processus automatisé modifié qui produirait le classement correct s'il devait être appliqué aux listes de mots des familles macro-jê et tupi en haut, et décrivez ce processus *brièvement*.

⚠ Cette tâche ne sera notée que dans le cas d'une égalité de points entre les meilleures équipes.

Les auteurs remercient Alejandra Vidal, Maria Konoshenko, Ilya Gruntov et Jamthô Suyá d'avoir répondu à leurs questions sur des langues spécifiques. —*Andreï Nikouline, Miléna Vénéva*

Rédacteurs : Ivan Derjanski (éditeur technique), Hugh Dobbs, Stanislav Gurévitch, Boris Iomdin, Liam McKnight, Andreï Nikouline (éditeur en chef), Aleksejs Peguševs, Jan Petr, Alexandre Piperski, Maria Rubinstein, Miléna Vénéva, Elysia Warner.

Texte français : Elysia Warner.

Bon courage !