

# 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 φ β f v   | K. | k g x ɣ q ɠ χ ɰ  | Y. | j ç (au début de la racine) | M. | m ɱ     |
| T. | t d ɗ θ ð ʈ ɖ   | R. | r r̥ ɽ ɺ ɻ ʂ ʃ ʄ | W. | w ɱ (au début de la racine) | N. | n ɲ ɳ ɰ |
| S. | s z ʃ ʒ ʂ ʐ ʑ ʒ   |    |                  |    |                             | Q. | ʈ ɖ     |
| H. | h ʕ ɦ ʕ̣ ʔ h ɦ ʔ, les voyelles, et j ç w ɱ (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 **jok:i** 'sel' de la langue kadiwéu est emprunté au mot guarani **juki**, et **?a:nj** 'année' du ipai (de Mesa Grande) est emprunté au mot espagnol **'ajno**.

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 vejoj.

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. ' = accent primaire, , = accent secondaire (plus faible que l'accent primaire), ◌: = son long, ◌◌ = son très court, X̂Y = X et Y sont prononcés comme un seul son, ◌́ = ton haut, ◌̀ = ton bas, ◌̂ = ton descendant, ◌̣ = son préglottalisé

(précédé par un bref blocage du passage d'air dans la gorge),  $\text{○}'$  = son éjectif (prononcé en bloquant brièvement le passage d'air dans la gorge),  $\text{○}$  = son sourd,  $\text{○}̃$  = son nasalisé (prononcé avec passage de l'air par le nez),  $\text{○}$  = voix craquée (tonalité basse et rauque),  $\text{ᵐ○}$  indique le flux d'air par le nez avant la consonne,  $\text{○}^h$  = consonne aspirée (prononcée avec une bouffée d'air),  $\text{○}^w$  = consonne labialisée (prononcée avec lèvres arrondies),  $\text{○}^j$  = son palatalisé (prononcé en dirigeant une partie de la langue près du palais dur).  $\text{a, æ, ε, i, ɪ, ɔ, ʊ, u, ə, ʌ, ɒ, ɐ, y, ø}$  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<sup>ère</sup> partie. La famille waykuruane (Argentine, Brésil, Paraguay)

|         | toba (de l'Est)        | pilagá                  | mocoví (de Chaco)             | kadiwéu                    |
|---------|------------------------|-------------------------|-------------------------------|----------------------------|
| nuage   | l=ʔok <sub>1</sub>     | 'lo=ʔok <sub>1</sub>    | naweyelek <sub>2</sub>        | lol:adi <sub>3</sub>       |
| feu     | nodek <sub>1</sub>     | 'd=oleʔ <sub>2</sub>    | norek <sub>1</sub>            | n=ol:edi <sub>2</sub>      |
| poisson | njaq <sub>1</sub>      | 'nijaq <sub>1</sub>     | naʎin <sub>2</sub>            | nij:ogo-dʒegi <sub>3</sub> |
| tête    | =qajk <sub>1</sub>     | =qajk <sub>1</sub>      | =qaik <sub>1</sub>            | =ak:ilo <sub>2</sub>       |
| tuer    | =alawat <sub>1</sub>   | =a'la:t <sub>1</sub>    | =alawat <sub>1</sub>          | =el:owadi <sub>1</sub>     |
| lune    | ʔawoʝok <sub>1</sub>   | ʔa'woʝok <sub>1</sub>   | ʃirajyo <sub>2</sub>          | ep:enaj <sub>3</sub>       |
| nez     | =mik <sub>1</sub>      | =mik <sub>1</sub>       | =mik <sub>1</sub>             | =m:iq:o <sub>1</sub>       |
| sel     | towe <sub>1</sub>      | ol'ʒek <sub>2</sub>     | ʔwe <sub>1</sub>              | jok:i <sub>1</sub>         |
| Pierre  | qaʔ <sub>1</sub>       | 'qaʔ <sub>1</sub>       | qaʔ <sub>1</sub>              | wet:iga <sub>2</sub>       |
| langue  | =atʃ-aʃat <sub>1</sub> | =a'tʃ-aʃat <sub>1</sub> | =oʔley-aʃan-aʃat <sub>2</sub> | =ok:el:i <sub>3</sub>      |

|             | algorithme A                      | algorithme B |  |
|-------------|-----------------------------------|--------------|--|
| manuel      | <p>distance lexicostatistique</p> |              | Indices de stabilité:<br>nuage 0,50<br>feu 0,50<br>poisson 0,50<br>tête 0,75<br>tuer 1,00<br>lune 0,50<br>nez 1,00<br>sel 0,67<br>pierre 0,75<br>langue 0,50 |
| automatique |                                   |              | Indices de stabilité:<br>nuage 0,50<br>feu 0,50<br>poisson 0,75<br>tête 0,75<br>tuer 1,00<br>lune 0,50<br>nez 1,00<br>sel 0,25<br>pierre 0,75<br>langue 0,50 |

II<sup>ème</sup> partie. La famille nubienne (Egypte, Soudan)

|         |           |           |         |           |           |            |
|---------|-----------|-----------|---------|-----------|-----------|------------|
|         | dongolawi | kenzi     | dilling | kadaru    | debri     | birgid     |
| tuer    | 'bɛ:₁     | be:₁      | hur₂    | wur-i₂    | wur-i₂    | fila:l-e₁  |
| lune    | u'n-at-t₁ | an-at-ti₁ | nɔn-ti₁ | nɔn-tu₁   | nɔn-to₁   | ma:l₂      |
| eau     | 'ɛss₁     | essi₁     | ɔti₁    | ɔto₁      | ɔtu₁      | ɛj₁        |
| donner  | 'tir₁     | tir₁      | ti₁     | ti₁       | ti₁       | te:r-n₁    |
| bon     | 'sere:₁   | sere:₁    | ken₂    | kɛn₂      | kɛŋ₂      | azze-n₃    |
| vent    | 'turug₁   | turug₁    | irf-i₂  | irf-o₂    | irf-o₂    | kurr-i₃    |
| cheveux | 'dil-ti₁  | si:r₂     | tel-ti₁ | til-tu₁   | til-tu₁   | ur=dill-e₁ |
| ventre  | 'tu:₁     | tu:₁      | te-te₂  | to₁       | to₁       | tu:₁       |
| dormir  | 'ne:r₁    | ne:r₁     | jer₁    | dwallɛli₂ | jer-i₁    | ne:r-i₁    |
| soleil  | 'masil₁   | masil₁    | ɛj₂     | aju₂      | ɛŋgal-to₃ | ʔi:zi₂     |

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

- (A) (2 points) La consonne **ɣ** 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<sup>ème</sup> partie. La famille mataguayo (Argentine, Bolivie, Paraguay)

|         | wichi (du Bas Bermejo)            | wichi (de Rivadavia)            | vejoz   | 'weenhayek   | iyojwa'aja'                           | manhui                                      | nivaklé (des Shichaam Lhavos) | nivaklé (des Chishamnee Lhavos) | maká                     |
|---------|-----------------------------------|---------------------------------|---|--|---------------------------------------|---|-------------------------------|---------------------------------|--------------------------|
| feu     | ʔitox <sub>1</sub>                | ʔitox <sub>1</sub>              | ʔitah <sub>1</sub>  | ʔi:tax <sub>1</sub>  | ʔh <sup>h</sup> wat <sub>2</sub>      | ʔe <sup>h</sup> it'e <sub>1</sub>           | ʔitax <sub>1</sub>            | ʔitax <sub>1</sub>              | feʔt <sub>2</sub>        |
| poisson | ʔwahat <sub>1</sub>               | wahat <sub>1</sub>              | wahat <sub>1</sub>  | ʔwa:hat <sub>1</sub>   | siʔjus <sub>-1</sub>                  | ʃiʔjus <sub>-1</sub>                        | saxetʃ <sub>-1</sub>          | saxetʃ <sub>-1</sub>            | sehets <sub>-1</sub>     |
| ped     | =patʃ <sup>h</sup> u <sub>1</sub> | =qol <sub>2</sub>               | =patʃ <sup>h</sup> o <sub>1</sub> ,<br>=kala <sub>2</sub> | =pa:kʔoʔ <sub>1</sub>  | =ʔsat <sub>3</sub>                    | =kaʔlaʔ <sub>2</sub>                        | =foʔ <sub>4</sub>             | =foʔ <sub>4</sub>               | =fʔiʔ <sub>5</sub>       |
| eau     | ʔinot <sub>1</sub>                | ʔinot <sub>1</sub>              | wah <sub>2</sub>  | ʔina:t <sub>1</sub>  | ʔi'nʔat <sub>1</sub>                  | ʔaʔnat <sub>1</sub>                         | jinaʔt <sub>1</sub>           | jinaʔt <sub>1</sub>             | iweliʔ <sub>3</sub>      |
| donner  | =ʔweŋ <sub>-u</sub> <sub>1</sub>  | =weŋ <sub>-u</sub> <sub>1</sub> | =ʔweŋ <sub>-o</sub> <sub>1</sub>                          | =ʔweŋ <sub>-o</sub> ʔ <sub>1</sub>                                   | =ʔweh <sup>h</sup> n-aʔm <sub>2</sub> | =ʔhajʔ <sub>3</sub> ,<br>=ʔweŋ <sub>2</sub> | =xut <sub>4</sub>             | =xut-ej <sub>4</sub>            | tis-ix <sub>5</sub>      |
| bon     | ʔis <sub>1</sub>                  | ʔis <sub>1</sub>                | ʔis <sub>1</sub>  | ʔis <sub>1</sub>   | ʔes <sub>1</sub>                      | ʔe <sup>h</sup> is <sub>1</sub>             | ʔis <sub>1</sub>              | ʔis <sub>1</sub>                | t=ejk'un-ej <sub>2</sub> |
| vent    | ʔinwok <sup>w</sup> <sub>1</sub>  | ʔinwək <sub>1</sub>             | ʔihwok <sup>w</sup> <sub>1</sub>                          | =ja:ʔ <sub>2</sub> ,<br>=x <sup>w</sup> ox <sup>w</sup> <sub>3</sub> | ʔhlah <sup>h</sup> wuʔ <sub>4</sub>   | ʔhlah <sup>h</sup> wuuʔ <sub>4</sub>        | ʔaβiʔm <sub>5</sub>           | ʔaβiʔm <sub>5</sub>             | t'unik'i <sub>6</sub>    |
| arbre   | haʔlo <sub>1</sub>                | hal <sub>2</sub>                | haʔla <sub>1</sub>  | haʔlaʔ <sub>1</sub>  | ʔaʔlaʔ <sub>1</sub>                   | ʔaʔla-k <sub>1</sub>                        | ʔaʔkxi-juk <sub>2</sub>       | jiʔklaʔ <sub>1</sub>            | naxka-k <sub>3</sub>     |
| cheveux | =ʔwule-j <sub>1</sub>             | =wule-j <sub>1</sub>            | =ʔwole-j <sub>1</sub>                                     | =ʔwo:le-ç <sub>1</sub> ,<br>hi:lenaç <sub>2</sub>                    | =ʔwole <sub>1</sub>                   | =ʔwole-j <sub>1</sub>                       | =ʔateʔtʃ <sub>3</sub>         | =jeʔs <sub>4</sub>              | =ʔewkux-its <sub>5</sub> |
| tuer    | =lon <sub>1</sub>                 | =lən <sub>1</sub>               | =lan <sub>1</sub>   | =la:ŋ <sub>1</sub>   | =ʔlaʔan <sub>1</sub>                  | =ʔlan <sub>1</sub>                          | =klan <sub>1</sub>            | =klan <sub>1</sub>              | =lan <sub>1</sub>        |

|             | algorithme A  | algorithme B  |   |
|-------------|---|---|---|
| manuel      | <p>0,11</p> <p>0,33</p> <p>0,44</p> <p>0,78</p> <p>0,80</p> <p>0,90</p> <p>wichi (du Bas Bermejo)<br/> wichi (de Rivadavia)<br/> vejoz<br/> 'weenhayek<br/> iyojwa'aja'<br/> manhui<br/> nivaklé (des Shichaam Lhavos)<br/> nivaklé (des Chishamnee Lhavos)<br/> maká</p>             | <p>0,13</p> <p>0,46</p> <p>0,61</p> <p>0,78</p> <p>0,83</p> <p>0,90</p> <p>wichi (du Bas Bermejo)<br/> wichi (de Rivadavia)<br/> vejoz<br/> 'weenhayek<br/> iyojwa'aja'<br/> manhui<br/> nivaklé (des Shichaam Lhavos)<br/> nivaklé (des Chishamnee Lhavos)<br/> maká</p>             | <p>Indices de stabilité :</p> <p>feu 0,78<br/> poisson 1,00<br/> pied 0,33<br/> eau 0,78<br/> donner 0,44<br/> bon 0,89<br/> vent 0,33<br/> arbre 0,78<br/> cheveux 0,67<br/> tuer 1,00</p> |
| automatique | <p>0,10</p> <p>0,20</p> <p>0,50</p> <p>0,70</p> <p>0,80</p> <p>0,80</p> <p>0,90</p> <p>wichi (du Bas Bermejo)<br/> wichi (de Rivadavia)<br/> vejoz<br/> 'weenhayek<br/> iyojwa'aja'<br/> manhui<br/> nivaklé (des Shichaam Lhavos)<br/> nivaklé (des Chishamnee Lhavos)<br/> maká</p> | <p>0,18</p> <p>0,28</p> <p>0,64</p> <p>0,80</p> <p>0,75</p> <p>0,80</p> <p>0,90</p> <p>wichi (du Bas Bermejo)<br/> wichi (de Rivadavia)<br/> vejoz<br/> 'weenhayek<br/> iyojwa'aja'<br/> manhui<br/> nivaklé (des Shichaam Lhavos)<br/> nivaklé (des Chishamnee Lhavos)<br/> maká</p> | <p>Indices de stabilité :</p> <p>feu 0,78<br/> poisson 0,44<br/> pied 0,33<br/> eau 0,56<br/> donner 0,67<br/> bon 0,89<br/> vent 0,22<br/> arbre 0,67<br/> cheveux 0,67<br/> tuer 1,00</p> |

**IV<sup>ème</sup> 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.

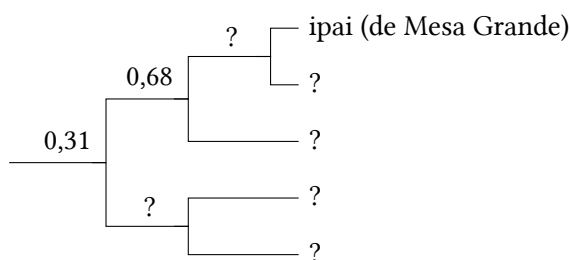
Pour vous aider, les deux indices de stabilité du mot 'tous' ont déjà été calculés. Dans un ordre aléatoire ils sont : 0,36 et 0,40.

|        | daur (de Hailar) | khamnigan (mand-chou) | bouriate (de Khori)   | nouveau-bargu | ööld    | khochoute | kalmouk         | khalkha               | ordos                          | shira yugur          | bonan    |
|--------|------------------|-----------------------|-----------------------|---------------|---------|-----------|-----------------|-----------------------|--------------------------------|----------------------|----------|
| tous   | hɔ:₁             | bölt₂                 | buxi:₃                | bygd₄         | tsug₅   | lug₅      | tsuk₅, xamak-₁  | pux₃, pugt₄, xamäg-₁  | pyyyte₄, xamuk <sup>h</sup> -₁ | tʃ <sup>h</sup> uq₅  | hanə-₂   |
| écorce | hails₁           | qalihön₁              | χoltöhön₂             | xalʃhu:₁      | xolts₂  | xalis₁    | dursn₃          | xəɮt <sup>h</sup> ɔs₂ | turusu₃                        | χalsən₁              | arasun₄  |
| ventre | ke:li₁           | getəhən₂              | gedehen₂              | gedy:₂        | ge:s₂   | gets₂     | gesn₂           | gitis₂, xiwɮij-₁      | ketysy₂                        | ketesən₂             | kele₁    |
| oiseau | dəgi-₁           | eiwan₁                | ʃubu:n₁               | ʃuwu:₁        | ʃuvu:₁  | ʃuwu:₁    | ʃowun₁          | ʃuwu₁                 | ʃuβu:₁                         | ʃu:n₁, peltʃər₂      | bendžer₂ |
| feu    | gali₁            | gal₁                  | gal₁                  | gal₁          | gal₁    | gal₁      | gal₁            | gal₁                  | qal₁                           | qal₁                 | χal₁     |
| chemin | terg-u:l₁        | qargūi₂               | χargi₂, zam-₁         | zam-₁         | dzam-₁  | dzam-₁    | xa:-lɤə₃        | tsam-₁                | tʃam-₁                         | mør₄                 | mor₄     |
| sel    | hata:₁           | dawhön₂               | dabhan₂               | dawuhu:₂      | daws₂   | daws₂     | dawsn₂          | tawsă₂                | taβusu₂                        | ta:psən₂             | dabsuŋ₂  |
| nager  | unpa-du₁         | umba-₁                | t <sup>h</sup> amar-₂ | umb-₁         | sele-₃  | umba-₁    | us-tɛi-₄, ø:m-₅ | siɮi-₃                | usu-tʃ <sup>h</sup> i-la-₄     | umpa-₁               | mba-₁    |
| eau    | ɔsɔ₁             | uxön₁                 | uhan₁                 | u:ha₁         | usn₁    | us₁       | usn₁            | usŏ₁                  | usun₁                          | q <sup>h</sup> usun₁ | sə₁      |
| vent   | kein₁            | halkin₂               | halxin₂               | halxi₂        | salxin₂ | salki₂    | saljkn₂         | saɮxi₂                | k <sup>h</sup> i:₁             | k <sup>h</sup> i:₁   | ki₁      |

**V<sup>ème</sup> 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 <sub>1</sub>                         | 'xɬ=ʔut <sub>2</sub>    | 'tʃkr=ot-i <sub>2</sub>                | lə=ʔuj <sub>1</sub>             | mə=put-k <sub>3</sub>               |
| oiseau | ʔitʃ=i=jer <sub>1</sub>                         | 'ʃa <sub>2</sub>        | 'ʔ=tʃ=sa <sub>2</sub>                  | aʔ=ʃa <sub>2</sub>              | ʔa:=ʃa:₂                            |
| os     | ʃn=a=s=ak <sub>1</sub>                          | 'ʃn=j=a:k <sub>1</sub>  | 'tʃ=j=a:k-a <sub>1</sub>               | 'ak <sub>1</sub>                | aq <sub>1</sub>                     |
| sec    | i=ro:-v-k <sub>1</sub>                          | 'ʃ=ʔar <sub>2</sub>     | 'ru-β-i <sub>1</sub>                   | 's=ʔa:j <sub>3</sub>            | sa:j <sub>3</sub>                   |
| viande | k <sup>w</sup> i:k <sup>w</sup> ay <sub>1</sub> | ʔi='ma:tʃ <sub>2</sub>  | 'k <sup>w</sup> e:=ʔo-β-a <sub>3</sub> | 'k <sup>w</sup> ak <sub>4</sub> | kuk <sup>w</sup> a:j-p <sub>1</sub> |
| cou    | maʃaqe <sub>1</sub>                             | 'm=puk <sub>2</sub>     | 'mlq <sub>1</sub>                      | i:=puk <sub>2</sub>             | i:=puk <sub>2</sub>                 |
| voir   | i=ju:-k <sub>1</sub>                            | 'wi:₂                   | 'ʔu:₁                                  | 'wi:w <sub>2</sub>              | ə=wu:w <sub>2</sub>                 |
| queue  | i:=ʔar <sub>1</sub>                             | 'ʃ=juʃ <sub>2</sub>     | 'β=hé <sub>3</sub>                     | ʃə=juʃ <sub>2</sub>             | xə=juʃ <sub>2</sub>                 |
| deux   | havik-k <sub>1</sub>                            | 'x=wak <sub>1</sub>     | 'h <sup>w</sup> âk-i <sub>1</sub>      | xə='wak <sub>1</sub>            | xə=wak <sub>1</sub>                 |
| année  | hu:ðe <sub>1</sub>                              | 'mat-'ka:m <sub>2</sub> | 'ʔ=tʃ <sup>h</sup> ur-a <sub>3</sub>   | mat-'wam <sub>2</sub>           | ʔa:n <sup>l</sup> -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.

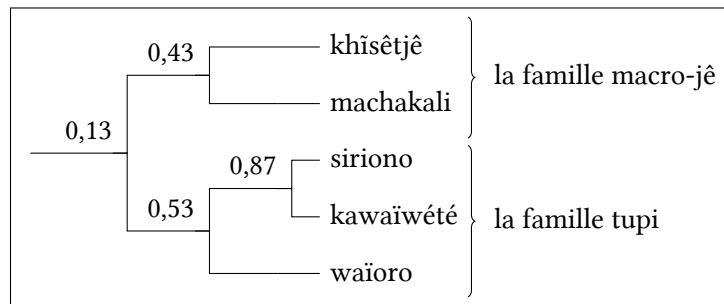
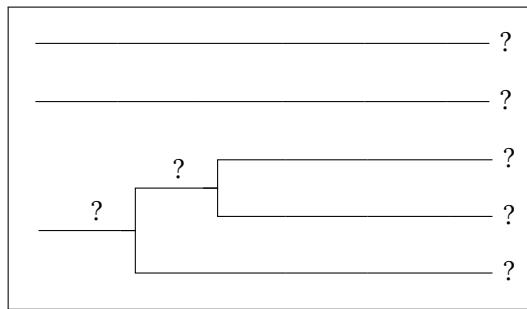
(K) (4 points) Expliquez la différence entre les algorithmes A et B.

**VI<sup>ème</sup> 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 apparentées. Examinez les listes de mots suivantes.

|         | A        | B                                  | Γ                        | Δ                 | E              |
|---------|----------|------------------------------------|--------------------------|-------------------|----------------|
| écorce  | e='e-ke  | h <sup>w</sup> i='k <sup>h</sup> Λ | kup='pe                  | mīβm='tɛaj        | = 'pe          |
| ventre  | 'e=rje   | = 't <sup>h</sup> igi              | =ā'ūn                    | = 'tæj            | =rɛ'wɛk        |
| sang    | e='ruki  | =ka' <sup>n</sup> ɾo               | =d̄z=a'u                 | = 'hɛβp           | =ru'i          |
| brûler  | = 'raī   | =rɔ='k <sup>h</sup> ɹɔ̃            | =po'k <sup>w</sup> a     | mū=...='haβp      | =ra'pi         |
| graisse | e='kira  | = 't <sup>h</sup> wəmi             | =d̄z=ap                  | = 'tuβp           | = 'kap         |
| pied    | 'e=i     | = 'h <sup>w</sup> aji              | = 'βi                    | =po'ta            | = 'pi          |
| main    | 'e=o     | =nī'k <sup>h</sup> ɹa              | = 'βo                    | = 'nīβm           | = 'pɔ          |
| lourd   | e='usi   | =wi't <sup>h</sup> ī               | =po'ti                   | =βp'təj           | =pɔ'ij         |
| foie    | 'e=ja    | = 'nba                             | =pi'a                    | =tɛiβpkī'nāj      | =pi'ʔa         |
| nouveau | e='jasu  | = 'ndiwi                           | =pa'gop                  | = 'tiβp           | =pia'u         |
| racine  | e='rao   | =ja'ɾe                             | kup=kujɔ'pe              | mīβm=nīβm=tɛa'tiə | =ra'pɔ         |
| peau    | 'e=i     | = 'k <sup>h</sup> Λ                | = 'pe                    | = 'tɛaj           | = 'pit         |
| queue   | e='rokoī | = 'nbi                             | =d̄z=o'k <sup>w</sup> aj | =nā:='kiβp        | = 'raj         |
| blanc   | 'e=ʃī    | =ja'k <sup>h</sup> a               | =d̄zi'ra                 | =βp'douɥ          | = 'sīɲ         |
| aile    | e='heo   | =ja'ɾa                             | =pe'o                    | =nī'māuɥ          | =pe'pɔ, =ji'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 intentionnellement 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 (khîsêtjê), 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 !