

## Kahdennekymmenennetensimmäiset kansainväliset kielitieteen olympialaiset

Brasília (Brasilia), 23.–31. heinäkuuta 2024

## Joukkuekilpailun tehtävä

Leksikostatistiikka on joukko menetelmiä, joiden tarkoituksena on arvioida, kuinka läheistä sukua kielet ovat toisilleen niiden sanastojen perusteella. Näitä menetelmiä sovelletaan tavallisesti pitkiin sanalistoihin, jotka asiantuntijat merkitsevät manuaalisesti ilmoittamalla, oletetaanko tietyn sanan olevan peräisin samasta lähteestä. Joskus kuitenkin kielitieteilijät käyttävät leksikostatistisia menetelmiä sanalistoihin, jotka on merkitty automaattisesti. Yksi tällainen automaattinen menettely perustuu *konsonanttiluokkien* käsitteeseen, jonka neuvostoliittolais-israelilainen kielitieteilijä Aharon Dolgopolsky esitti vuonna 1964.

|    |   |    |                    |    |                     |    |          |
|----|---|----|--------------------|----|---------------------|----|----------|
| P. | p b β φ β f v   | K. | k g x γ q ɣ χ u    | Y. | j ç (juuren alussa) | M. | m ɱ      |
| T. | t d d̥ θ ð t̥ d̥  | R. | r r̥ ɾ ɹ l l̥ ʎ ʎ̥ | W. | w ɱ (juuren alussa) | N. | n ɲ ɳ ɳ̥ |
| S. | s z ʃ ʒ ʂ ʐ z̥ c ɟ  |    |                    |    |                     | Q. | ṯ ḏ      |
| H. | ħ ʕ ɦ ʔ ʕ h ɦ ʔ, vokaalit ja j ç w ɱ (paitsi juuren alussa) |    |                    |    |                     |    |          |

## Dolgopolskyn konsonanttiluokat

Alla on merkittyjen sanalistojen katkelmia useista maailman kielikunnista. Merkinnät annetaan alaindeksein. Näiden listojen perusteella on luotu kielten sukupuut käyttäen kahta yksinkertaistettua versiota ns. *StarlingNj*-algoritmista, minkä lisäksi jokaiselle sanalle on annettu stabiilisuusindeksi. Yläosan puut ja stabiilisuusindeksit perustuvat manuaalisesti merkittyihin sanalistoihin ja alaosan puut ja stabiilisuusindeksit sanalistoihin, jotka on merkitty automaattisesti. Jokaisen sanalistan perusteella on rakennettu kaksi puuta, jotka seuraavat kahta algoritmiversiota: Algoritmi A ja Algoritmi B. Huomatkaa, että joissakin tapauksissa on useita mahdollisia puita, jotka vastaavat yhtä sanalista; näissä tapauksissa vain yksi puu valittiin satunnaisesti. Jokaisen puun jokaiselle noodille (solmulle) on määriteltä leksikostatistinen etäisyys. Mitä suurempi etäisyys, sitä läheisempi on kielten välinen suhde. Tarkempi termi olisi siis ”käänteinen leksikostatistinen etäisyys”, ei ”leksikostatistinen etäisyys”. Yksinkertaisuuden vuoksi käytämme tässä tehtävässä termiä ”leksikostatistinen etäisyys”.

Sekä stabiilisuusindeksit että leksikostatistiset etäisyydet pyöristetään kahteen desimaaliin. Jos kolmas numero desimaalipilkun jälkeen on pienempi kuin 5, arvo pyöristetään alaspäin; muuten se pyöristetään ylöspäin. Esimerkiksi, 2,836 pyöristetään 2,84:ään, 0,705 pyöristetään 0,71:een, ja 0,703 pyöristetään 0,70:een. Pyöristäminen pätee vain arvoihin, jotka näytetään ihmislukijoille. Toisin sanoen, tietokone, joka käyttää algoritmeja, ”näkee” pyöristämättömät arvot.

Huomatkaa, että joidenkin sanojen tiedetään tai epäillään olevan lainattu muista kielistä. Esimerkiksi kadiveun kielen sana **jok:i** ’suola’ on lainattu guaranin kielen sanasta **juki** ja ’iipain kielen (Mesa Granden murteen) sana **ʔa:nj** ’vuosi’ on lainattu espanjan kielen sanasta **’ajno**.

Joissakin tapauksissa sanalistoissa annetaan yhdelle merkitykselle useita synonyymejä, jotka erotetaan toisistaan pilkulla. Esimerkkinä mainittakoon vehvosin kielen ’jalkaa’ tarkoittavat sanat.

Alla olevissa tiedoissa kaikki prefiksit on erotettu ”=”-merkillä ja kaikki suffiksit ”-”-merkillä. Joitain sanoja käytetään vain prefiksien kanssa. Nämä alkavat ”=”-merkillä.

Tiedot on kirjoitettu kansainvälisiä foneettisia aakkosia käyttäen. <sup>ˈ</sup> = pääpaino, <sub>ˈ</sub> = sivupaino (pääpainoa heikompi), ː = pitkä äänne, ̥ = ylilyhyt äänne, X̄Y = X ja Y äännetään yhtenä ääntenä, ˆ = korkea tooni, ˆ = matala tooni, ˆ = laskeva tooni, ˆ = preglottalisoitu äänne (jota edeltää lyhyt ilma-  
virran estyminen kurkussa), ˆ = ejetiivi (konsonantti, jossa kurkunpää suljetaan ja sitten avataan),

◌̥ = soinniton konsonantti, ◌̃ = nasalisoitu äänne (nenän kautta lausuttava äänne), ◌̚ = narinaääni, <sup>n</sup>◌ osoittaa ilmapvirtaa nenän kautta konsonantin edellä, ◌<sup>h</sup> = aspiroitu (ilmaa puhaltaen lausuttava) konsonantti, ◌<sup>w</sup> = labialisoitu (pyöristetyillä huulilla lausuttava) konsonantti, ◌<sup>j</sup> = palatalisoitu (kielen keskiosaa kovaan kitalakeen lähestyen lausuttava) konsonantti. **ɑ, æ, ɛ, ɪ, i, ɔ, ʊ, ʉ, ɐ, ʌ, ɒ, ɘ, ɤ, ɞ, ø** ovat vokaaleja. Muut erikoismerkit ovat konsonantteja.

⚠ Minkään tehtävässä mainitun kielen osaamisesta ei ole apua tehtävää ratkaistaessa.

### I. osa. Guaikuru-kielet (Argentiina, Brasilia, Paraguay)

|        | toba (itämurre)        | pilagá                  | mocoví (Chacon murre)         | kadiveu                    |
|--------|------------------------|-------------------------|-------------------------------|----------------------------|
| pilvi  | l=ʔok <sub>1</sub>     | 'lo=ʔok <sub>1</sub>    | naweyelek <sub>2</sub>        | lol:adi <sub>3</sub>       |
| tuli   | nodek <sub>1</sub>     | 'd=oleʔ <sub>2</sub>    | norek <sub>1</sub>            | n=ol:edi <sub>2</sub>      |
| kala   | njaq <sub>1</sub>      | 'nijaq <sub>1</sub>     | naʎin <sub>2</sub>            | nij:ogo-ḏzegi <sub>3</sub> |
| pää    | =qajk <sub>1</sub>     | =ʔajk <sub>1</sub>      | =qaik <sub>1</sub>            | =ak:ilo <sub>2</sub>       |
| tappaa | =alawat <sub>1</sub>   | =a'la:t <sub>1</sub>    | =alawat <sub>1</sub>          | =el:owadi <sub>1</sub>     |
| kuu    | ʔawoɔɔjk <sub>1</sub>  | ʔa'woʔɔjk <sub>1</sub>  | ʃirajɔ <sub>2</sub>           | ep:enaj <sub>3</sub>       |
| nenä   | =mik <sub>1</sub>      | =ʔmik <sub>1</sub>      | =mik <sub>1</sub>             | =m:iq:o <sub>1</sub>       |
| suola  | towe <sub>1</sub>      | ol'ɣek <sub>2</sub>     | ʔwe <sub>1</sub>              | jok:i <sub>1</sub>         |
| kivi   | qaʔ <sub>1</sub>       | 'qaʔ <sub>1</sub>       | qaʔ <sub>1</sub>              | wet:iɡa <sub>2</sub>       |
| kieli  | =atʃ-aʃat <sub>1</sub> | =a'tʃ-aʃat <sub>1</sub> | =oʔley-aʃan-aʃat <sub>2</sub> | =ok:el:i <sub>3</sub>      |

|                 | A-algoritmi                        | B-algoritmi |  |
|-----------------|------------------------------------|-------------|--|
| manuaalinen     | <p>leksikostatistinen etäisyys</p> |             | Stabiilisuusindeksit:<br>pilvi 0,50<br>tuli 0,50<br>kala 0,50<br>pää 0,75<br>tappaa 1,00<br>kuu 0,50<br>nenä 1,00<br>suola 0,67<br>kivi 0,75<br>kieli 0,50 |
| automaattisoitu |                                    |             | Stabiilisuusindeksit:<br>pilvi 0,50<br>tuli 0,50<br>kala 0,75<br>pää 0,75<br>tappaa 1,00<br>kuu 0,50<br>nenä 1,00<br>suola 0,25<br>kivi 0,75<br>kieli 0,50 |

## II. osa. Nubialainen kielikunta (Egypti, Sudan)

|         |           |           |         |           |           |            |
|---------|-----------|-----------|---------|-----------|-----------|------------|
|         | dongolau  | kenuzi    | dilling | kadaru    | debri     | birgid     |
| tappaa  | 'bɛ:₁     | be:₁      | hur₂    | wur-i₂    | wur-i₂    | fila:l-e₁  |
| kuu     | u'n-at-t₁ | an-at-ti₁ | nɔn-ti₁ | nɔn-tu₁   | nɔn-to₁   | ma:l₂      |
| vesi    | 'ɛss₁     | essi₁     | ɔti₁    | ɔto₁      | ɔtu₁      | eji₁       |
| antaa   | 'tir₁     | tir₁      | ti₁     | ti₁       | ti₁       | te:-n₁     |
| hyvä    | 'sɛrɛ:₁   | sere:₁    | ken₂    | kɛn₂      | kɛŋ₂      | azze-n₃    |
| tuuli   | 'turug₁   | turug₁    | irf-i₂  | irf-o₂    | irf-o₂    | kurr-i₃    |
| tukka   | 'dil-ti₁  | si:r₂     | tɛl-ti₁ | til-tu₁   | til-tu₁   | ur=dill-e₁ |
| vatsa   | 'tu:₁     | tu:₁      | te-te₂  | to₁       | to₁       | tu:₁       |
| nukkua  | 'nɛ:r₁    | ne:r₁     | jer₁    | dwallɛli₂ | jer-i₁    | ne:r-i₁    |
| aurinko | 'masil₁   | masil₁    | ɛj₂     | aju₂      | ɛŋgal-to₃ | ʔi:zi₂     |

|                 | A-algoritmi | B-algoritmi |   |
|-----------------|-------------|-------------|---|
| manuaalinen     |             |             | Stabiilisuusindeksit:<br>tappaa 0,50<br>kuu 0,83<br>vesi 1,00<br>antaa 1,00<br>hyvä 0,50<br>tuuli 0,50<br>tukka 0,83<br>vatsa 0,83<br>nukkua 0,83<br>aurinko 0,50 |
| automaattisoitu |             |             | Stabiilisuusindeksit:<br>tappaa 0,33<br>kuu 0,50<br>vesi 0,50<br>antaa 0,67<br>hyvä 0,50<br>tuuli 0,50<br>tukka 0,83<br>vatsa 1,00<br>nukkua 0,50<br>aurinko 0,50 |

- (A) (2 pistettä)  $\mathfrak{R}$  lausutaan kuin ranskan  $r$ , kielen takaosassa. Mihin Dolgopolsky-luokkaan se kuuluu ja miten saitte sen selville?
- (B) (2 pistettä) Vasemmassa yläkulmassa oleva nubialaisten kielten puu on vain yksi kahdesta mahdollisesta puusta tässä algoritmin ja merkintätyyppin yhdistelmässä. Piirtäkää toinen mahdollinen puu.
- (C) (2 pistettä) Vasemmassa alakulmassa oleva nubialaisten kielten puu on vain yksi kahdesta mahdollisesta puusta tässä algoritmin ja merkintätyyppin yhdistelmässä. Piirtäkää toinen mahdollinen puu.
- (D) (2 pistettä) Leksikostatistinen etäisyys 0,49 (joka on määritetty nubialaisten kielten oikeassa yläkulmassa olevan puun juurelle) on pyöristetty kahteen desimaaliin, kuten jotkut muutkin etäisyydet tässä tehtävässä. Mikä on tarkka etäisyys?

### III. osa. Mataguaiolainen kielikunta (Argentiina, Bolivia, Paraguay)

|        | vitši (Bermejon alajuoksun murre) | vitši (Rivadavian murre) | vehvos                                      | veehnajekki  | ijohvaaha               | manhui                                     | nivakle (alajuoksun murre) | nivakle (yläjuoksun murre) | maká                     |
|--------|-----------------------------------|--------------------------|---|--|-------------------------|--|----------------------------|----------------------------|--------------------------|
| tuli   | ʔitox <sub>1</sub>                | ʔitox <sub>1</sub>       | ʔitah <sub>1</sub>                          | ʔi:tax <sub>1</sub>  | ʔhwat <sub>2</sub>      | ʔeite <sub>1</sub>                         | ʔitax <sub>1</sub>         | ʔitax <sub>1</sub>         | feʔt <sub>2</sub>        |
| kala   | 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>     |
| jalka  | =patʃ <sub>u1</sub>               | =qol <sub>2</sub>        | =patʃ <sub>o1</sub> ,<br>=kala <sub>2</sub> | =pa:kʔ <sub>o1</sub>   | =ʔsat <sub>3</sub>      | =kaʔla <sub>2</sub>                        | =foʔ <sub>4</sub>          | =foʔ <sub>4</sub>          | =fʔi <sub>5</sub>        |
| vesi   | ʔinot <sub>1</sub>                | ʔinot <sub>1</sub>       | wah <sub>2</sub>                            | ʔina:t <sub>1</sub>  | ʔiʔnat <sub>1</sub>     | ʔaʔnat <sub>1</sub>                        | jinaʔt <sub>1</sub>        | jinaʔt <sub>1</sub>        | iweli <sub>3</sub>       |
| antaa  | =ʔwen <sub>u1</sub>               | =wen <sub>u1</sub>       | =ʔwen <sub>o1</sub>                         | =ʔwen <sub>o1</sub>  | =ʔwehn-aʔm <sub>2</sub> | =ʔhaj <sub>3</sub> ,<br>=ʔwen <sub>2</sub> | =xut <sub>4</sub>          | =xut-ej <sub>4</sub>       | tis-ix <sub>5</sub>      |
| hyvä   | ʔis <sub>1</sub>                  | ʔis <sub>1</sub>         | ʔis <sub>1</sub>                            | ʔis <sub>1</sub>   | ʔes <sub>1</sub>        | ʔeis <sub>1</sub>                          | ʔis <sub>1</sub>           | ʔis <sub>1</sub>           | t=ejkʔun-ej <sub>2</sub> |
| tuuli  | ʔinwok <sup>w</sup> <sub>1</sub>  | ʔinwok <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> | ʔhlahwu <sub>4</sub>    | ʔhlahwu <sub>4</sub>                       | ʔaβiʔm <sub>5</sub>        | ʔaβiʔm <sub>5</sub>        | tʔunikʔi <sub>6</sub>    |
| puu    | haʔlo <sub>1</sub>                | hal <sub>1</sub>         | haʔla <sub>1</sub>                          | haʔla <sub>1</sub>   | ʔaʔla <sub>1</sub>      | ʔaʔla-k <sub>1</sub>                       | ʔaʔkxi-juk <sub>2</sub>    | jiʔkla <sub>1</sub>        | naxka-k <sub>3</sub>     |
| tukka  | =ʔwule-j <sub>1</sub>             | =wule-j <sub>1</sub>     | =ʔwole-j <sub>1</sub>                       | =ʔwo:le-ç <sub>1</sub> ,<br>hi:lenax <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> |
| tappaa | =lon <sub>1</sub>                 | =lon <sub>1</sub>        | =lan <sub>1</sub>                           | =la:ŋ <sub>1</sub>   | =ʔlaʔan <sub>1</sub>    | =ʔlan <sub>1</sub>                         | =kla <sub>1</sub>          | =kla <sub>1</sub>          | =lan <sub>1</sub>        |

|               | A-algoritmi | B-algoritmi |  |
|---------------|-------------|-------------|--|
| manuaalinen   |             |             | Stabiilisuusindeksit:<br>tuli 0,78<br>kala 1,00<br>jalka 0,33<br>vesi 0,78<br>antaa 0,44<br>hyvä 0,89<br>tuuli 0,33<br>puu 0,78<br>tukka 0,67<br>tappaa 1,00 |
| automaatioitu |             |             | Stabiilisuusindeksit:<br>tuli 0,78<br>kala 0,44<br>jalka 0,33<br>vesi 0,56<br>antaa 0,67<br>hyvä 0,89<br>tuuli 0,22<br>puu 0,67<br>tukka 0,67<br>tappaa 1,00 |

#### IV. osa. Mongolilainen kielikunta (Kiinan kansantasavalta, Mongolia, Venäjä)

(E) (10 pistettä) Tarkastelkaa seuraavaa sanalista. Laskekaa sekä manuaalisia että automatisoituja merkintöjä vastaavat stabiilisuusindeksit.

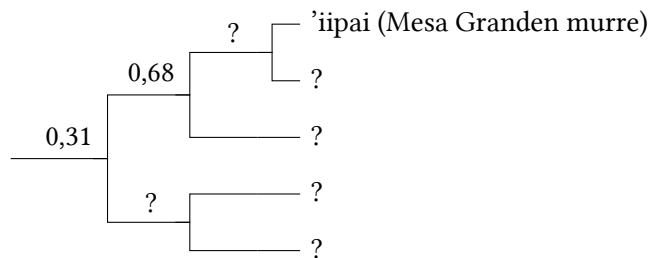
Auttaaksemme teitä, olemme jo laskeneet molemmat stabiilisuusindeksit 'kaikki'-sanalle. Satunnaisessa järjestyksessä nämä ovat 0,36 ja 0,40.

|        | daguuri<br>(hailarin<br>murre) | hamnigani<br>(mantšun<br>murre) | burjaatti<br>(horin<br>murre) | uus-<br>barguutti | ööldi    | hošuutti | kalmukki           | halha                      | ordos                | šira-<br>juguuri   | baoan    |
|--------|--------------------------------|---------------------------------|-------------------------------|-------------------|----------|----------|--------------------|----------------------------|----------------------|--------------------|----------|
| kaikki | hɔ:₁                           | bɔlt₂                           | bɔxi:₃                        | bygd₄             | tsug₅    | lug₅     | tsuk₅,<br>xamak₋₁  | pux₃,<br>pugt₄,<br>xamăg₋₁ | pyyyte₄,<br>xamukʰ₋₁ | tʃʰuq₅             | hanə₂    |
| kaarna | hails₁                         | qalihon₁                        | χoltɔhɔn₂                     | xalʰhu:₁          | xolts₂   | xalis₁   | dursn₃             | xɔɣtʰɔs₂                   | turusu₃              | χalsən₁            | arasun₄  |
| vatsa  | ke:li₁                         | gɔtɔhɔn₂                        | gedehen₂                      | gedy:₂            | ge:s₂    | gets₂    | gesn₂              | gitis₂,<br>xiwɣij₋₁        | ketysy₂              | ketesən₂           | kele₁    |
| lintu  | dəgi₋₁                         | eiwan₁                          | ʃubu:n₁                       | ʃuwu:₁            | ʃuvu:₁   | ʃuwu:₁   | ʃowun₁             | ʃuwu₁                      | ʃuβu:₁               | ʃu:n₁,<br>peltʃɔr₂ | bendzər₂ |
| tuli   | gali₁                          | gal₁                            | gal₁                          | gal₁              | gal₁     | gal₁     | gal₁               | gal₁                       | qal₁                 | qal₁               | χal₁     |
| tie    | terg-u:l₁                      | qargɔi₂                         | χargi₂,<br>zam₋₁              | zam₋₁             | dzam₋₁   | dzam₋₁   | xa:-lɔə₃           | tsam₋₁                     | tʃam₋₁               | mør₄               | mor₄     |
| suola  | hata:₁                         | dawhɔn₂                         | dabhan₂                       | dawuhu:₂          | daws₂    | daws₂    | dawsn₂             | tawsă₂                     | taβusu₂              | ta:psən₂           | dabsuŋ₂  |
| uida   | unpa-du₁                       | ɔmba₋₁                          | tʰamar₋₂                      | umb₋₁             | sele₋₃   | umba₋₁   | us-tɛi₋₄,<br>ɔ:m₋₅ | siɣi₋₃                     | usu-tʃʰi-la₋₄        | umpa₋₁             | mba₋₁    |
| vesi   | ɔsɔ₁                           | ɔxɔn₁                           | uhan₁                         | u:ha₁             | usn₁     | us₁      | usn₁               | ɔsɔ₁                       | usun₁                | qʰusun₁            | sə₁      |
| tuuli  | kein₁                          | halkin₂                         | halxin₂                       | halxi₂            | salʰxin₂ | salkʲi₂  | salʲkn₂            | salɣxi₂                    | kʰi:₁                | kʰi:₁              | ki₁      |

**V. osa. Jumalainen kielikunta** (Meksiko, Yhdysvallat)

(F) (8 pistettä) Tarkastelkaa seuraavaa sanalista. Alla on puu, joka on rakennettu saman sanalistan perusteella. Jotkut tiedot (kielten nimet ja leksikostatistiset etäisyydet) puuttuvat. Täydentäkää aukot. Ilmoittakaa, onko puu manuaalinen vai automaattinen, sekä onko se luotu algoritmilla A vai B.

|       | mohave  | kokopa                  | javapai                                 | tiipai (Jamulin murre)          | 'iipai (Mesa Granden murre)         |
|-------|---|-------------------------|---|---------------------------------|-------------------------------------|
| lyhyt | wena=wen-a <sub>1</sub>                         | 'xʌ=ʔut <sub>2</sub>    | 'tʃkr=ot-i <sub>2</sub>                 | lə='ʔuŋ <sub>1</sub>            | mə=put-k <sub>3</sub>               |
| lintu | ʔitʃ=i=jer <sub>1</sub>                         | 'ʃa <sub>2</sub>        | 'ʔ=tʃ=sa <sub>2</sub>                   | aʔ='ʃa <sub>2</sub>             | ʔa:=ʃa:₂                            |
| luu   | ŋ=a=s=ak <sub>1</sub>                           | 'ŋ=j=a:k <sub>1</sub>   | 'tʃ=j=a:k-a <sub>1</sub>                | 'ak <sub>1</sub>                | aq <sub>1</sub>                     |
| kuiva | 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>                   |
| liha  | k <sup>w</sup> i:k <sup>w</sup> ay <sub>1</sub> | ʔi='ma:tʃ <sub>2</sub>  | 'k <sup>w</sup> e:='θo-β-a <sub>3</sub> | 'k <sup>w</sup> ak <sub>4</sub> | kuk <sup>w</sup> a:j-p <sub>1</sub> |
| kaula | maʎaqe <sub>1</sub>                             | 'm=puk <sub>2</sub>     | 'mlq <sub>1</sub>                       | i:='puk <sub>2</sub>            | i:=puk <sub>2</sub>                 |
| nähdä | i=ju:-k <sub>1</sub>                            | 'wi:₂                   | 'ʔu:₁                                   | 'wi:w <sub>2</sub>              | ə=wu:w <sub>2</sub>                 |
| häntä | i:=ʔar <sub>1</sub>                             | 'ʃ=juʎ <sub>2</sub>     | 'β=hé <sub>3</sub>                      | ʃə='juʎ <sub>2</sub>            | xə=juʎ <sub>2</sub>                 |
| kaksi | havik-k <sub>1</sub>                            | 'x=wak <sub>1</sub>     | 'h <sup>w</sup> ák-i <sub>1</sub>       | xə='wak <sub>1</sub>            | xə=wak <sub>1</sub>                 |
| vuosi | 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>i</sup> -₁                |



(G) (20 pistettä) On luotu jotkut muut jumalaisten kielten puut, joita vastaavat seuraavat puun juuressa (jokaisesta puusta kauimpana vasemmalla) olevat leksikostatistiset etäisyydet:

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

Piirtäkää jokainen näistä puista. Ilmoittakaa kunkin puun osalta, onko se manuaalinen vai automaattinen, sekä onko se luotu algoritmilla A vai B.

(H) (3 pistettä) Kaksi (G)-alitehtävässä luetelluista etäisyyksistä on pyöristetty kahteen desimaaliin: 0,23 on pyöristetty 0,225:stä. Mikä muu etäisyys on pyöristetty ja mikä on sen tarkka arvo?

(I) (4 pistettä) Selittäkää, miten stabiilisuusindeksit lasketaan.

(J) (5 pistettä) Selittäkää, miten leksikostatistiset etäisyydet lasketaan.

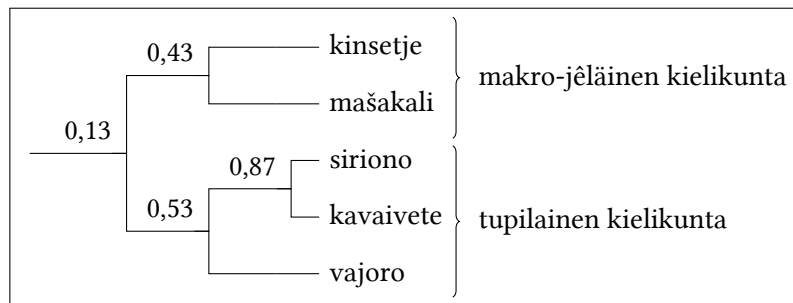
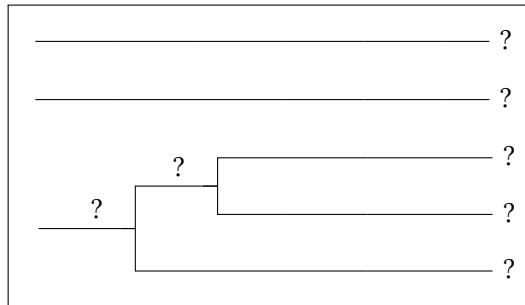
(K) (4 pistettä) Selittäkää A- ja B-algoritmien välinen ero.

## VI. osa. Makro-jêläinen kielikunta ja tupilainen kielikunta (Brasilia, Bolivia)

(L) (28 pistettä) Makro-jêläiset ja tupilaiset kielet muodostavat kaksi suurta kielikuntaa, joihin kuuluvia kieliä puhutaan Etelä-Amerikassa. Jotkut kielitieteilijät uskovat niiden olevan kaukaisia sukulaisia. Tarkastelkaa seuraavia sanalistoja.

|           | A        | B                                  | Γ                        | Δ                 | E              |
|-----------|----------|------------------------------------|--------------------------|-------------------|----------------|
| kaarna    | e='e-ke  | h <sup>w</sup> ĩ='k <sup>h</sup> Λ | kup='pe                  | mĩβm='tɛaj        | = 'pe          |
| vatsa     | 'e=rje   | = 't <sup>h</sup> igi              | =ã'ũn                    | = 'tɛj            | =rɛ'wɛk        |
| veri      | e='ruki  | =ka' <sup>n</sup> bɾo              | =d̄z=a'ɛ                 | = 'hɛβp           | =ru'i          |
| polttaa   | = 'rai   | =rɔ='k <sup>h</sup> ɹã             | =po'k <sup>w</sup> a     | mũ=...='haβp      | =ra'pi         |
| rasva     | e='kira  | = 't <sup>h</sup> wɔmi             | = 'd̄z=ap                | = 'tuβp           | = 'kap         |
| jalka     | 'e=i     | = 'h <sup>w</sup> aji              | = 'βi                    | =po'ta            | = 'pi          |
| käsi      | 'e=o     | =nĩ'k <sup>h</sup> ɹa              | = 'βo                    | = 'nĩβm           | = 'pɔ          |
| raskas    | e='usi   | =wi't <sup>h</sup> ĩ               | =po'ti                   | =βp'tɔj           | =pɔ'ij         |
| maksa     | 'e=ja    | = 'nba                             | =pi'a                    | =tɛiβpkĩ'nãj      | =pi'ʔa         |
| uusi      | e='jasu  | = 'ndiwi                           | =pa'gop                  | = 'tiβp           | =pia'u         |
| juuri     | e='rao   | =ja'ɾe                             | kup=kujɔ'pe              | mĩβm=nĩβm=tɛa'tiã | =ra'pɔ         |
| iho       | 'e=i     | = 'k <sup>h</sup> Λ                | = 'pe                    | = 'tɛaj           | = 'pit         |
| häntä     | e='rokoi | = 'nbi                             | =d̄z=o'k <sup>w</sup> aj | =nã:='kiβp        | = 'raj         |
| valkoinen | 'e=fĩ    | =ja'k <sup>h</sup> a               | =d̄zi'ra                 | =βp'douɥ          | = 'sĩŋ         |
| siipi     | e='heo   | =ja'ɾa                             | =pe'o                    | =nĩ'mãuɥ          | =pe'pɔ, =ji'wa |

Alla on kaksi puuta, jotka on rakennettu samojen sanalistojen perusteella. Jotkut tiedot (kielten nimet ja leksikostatistiset etäisyydet) puuttuvat. Täydentäkää aukot. Ilmoittakaa kunkin puun osalta, onko se manuaalinen vai automaattinen, sekä onko se luotu algoritmilla A vai B.



| A | B | Γ | Δ | E |
|---|---|---|---|---|
| ? | ? | ? | ? | ? |

⚠ Manuaaliset merkinnät ja stabiilisuusindeksit jätettiin tarkoituksella pois tästä alitehtävästä.

(M) (10 pistettä) Dolgopolsky-luokkiin perustuvat automaattiset menettelyt voivat tuottaa väärää tuloksia. Tässä esimerkissä automaattinen menettely havaitsee enemmän yhtäläisyyksiä sironon kielen ja tietyn makro-jäläisen kielen (kinsetjen kielen) välillä kuin sironon ja muiden tupilaisten kielten välillä. Ehdottakaa muokattua automaattista menettelyä, joka tuottaisi oikean makro-jäläisten ja tupilaisten kielten luokittelun, jos sitä sovellettaisiin yllä oleviin sanalistoihin, ja kuvatkaa tämä *lyhyesti*.

△ Tämä alitehtävä arvostellaan ainoastaan siinä tapauksessa, että parhaat pisteet saavuttaneet joukkueet päätyvät tasapisteisiin.

Kiitokset Alejandra Vidalille, Maria Konoshenkolle, Ilja Gruntoville ja Jamthô Suyälle muutamia kieliä koskeviin kysymyksiin vastaamisesta. —*Andrei Nikulin, Milena Veneva*

---

**Toimittajat:** Ivan Deržanski (tekn. toim.), Hugh Dobbs, Stanislav Gurevitš, Boris Iomdin, Eimear McKnight, Andrei Nikulin (päätoimittaja), Aleksejs Peguševs, Jan Petr, Aleksandr Piperski, Marija Rubinštein, Milena Veneva, Elysia Warner.

**Suomenkielinen teksti:** Andrei Nikulin, Seppo Kittilä.

Onnea kilpailuun!