nl

Eenentwintigste Internationale Taalkunde-Olympiade

Brasilia (Brazilië), 23–31 juli 2024

Opgave van de groepswedstrijd

Lexicostatistiek is een groep van methoden ontworpen om een inschatting te maken van hoe nauw verwant verschillende talen zijn op basis van hun vocabulaire. Deze methoden worden gewoonlijk toegepast op lange woordenlijsten door deskundigen, die handmatige annotaties toevoegen en aangeven of een bepaald woordpaar uit dezelfde bron afkomstig is. Soms passen taalkundigen echter lexicostatistische methoden toe op woordenlijsten die worden geannoteerd door geautomatiseerde procedures. Eén zo'n procedure is gebaseerd op het concept van de *medeklinkerklassen*, geïntroduceerd in 1964 door de Sovjet-Israëlische taalkundige Aharon Dolgopolsky.

P.	p b ɓ ɸ β f v	K.	k g x ɣ q ɢ χ ɯ	Y.	j ç (aan het begin van de wortel)	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 (aan het begin van de wortel)	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 ɦ ʔ, klinkers en j ç w ʍ (behalve aan het begin van de wortel)						

Medeklinkerklassen van Dolgopolsky

Hieronder volgen geannoteerde fragmenten van woordenlijsten van meerdere taalfamilies van de wereld. De annotaties worden gegeven als cijfers in subscript. Op basis van deze lijsten zijn taalfamiliestambomen geconstrueerd met behulp van twee vereenvoudigde versies van het zogenaamde *Starling-NJ*-algoritme, en is aan elk woord een *stabiliteitsindex* toegewezen. De bomen en stabiliteitsindices bovenaan zijn gebaseerd op handmatig geannoteerde lijsten, en die onderaan zijn gebaseerd op automatisch geannoteerde lijsten. Voor elke woordenlijst zijn er twee bomen, geconstrueerd met twee versies van het algoritme: Algoritme A en Algoritme B. Merk op dat er per woordenlijst soms meerdere bomen mogelijk zijn; in dergelijke gevallen werd één boom willekeurig geselecteerd. Aan elk knooppunt in elke boom is een lexicostatistische afstand toegewezen. Hoe groter de afstand, hoe nauwer de relatie tussen de talen. De preciezere term is dus "omgekeerde lexicostatistische afstand" eerder dan "lexicostatistische afstand". Omwille van de eenvoudigheid wordt in deze opgave de term "lexicostatistische afstand" gebruikt.

Zowel de stabiliteitsindices als de lexicostatistische afstanden zijn afgerond op twee decimalen. Als het derde cijfer achter de komma kleiner is dan 5, rond dan naar beneden af; rond als dit niet het geval is naar boven af. Bijvoorbeeld: 2,836 wordt op 2,84 afgerond, 0,705 op 0,71 en 0,703 op 0,70. De afronding is alleen van toepassing op de waarden die aan menselijke lezers worden getoond. Met andere woorden, de computer waarop de algoritmen lopen, "ziet" de niet-afgeronde waarden.

Merk op dat het bekend is of vermoed wordt dat enkele woorden zijn ontleend aan andere talen. Zo is **jokːi** 'zout' in de Kadiwéu-taal ontleend aan het Guaraní **juki**, en is **ʔaːnʲ** 'jaar' in het Ipai (Mesa Grande) ontleend aan het Spaanse **ˈaɲo**.

In sommige gevallen worden meerdere synoniemen van een enkele betekenis gegeven in de woordenlijsten, gescheiden door komma's. Een voorbeeld is 'voet' in het Vejoz.

In de gegevens hieronder worden alle voorvoegsels gescheiden door een "=" teken en alle achtervoegsels door een "-" teken. Sommige woorden komen alleen voor met voorvoegsels. Deze woorden beginnen met een "="-teken.

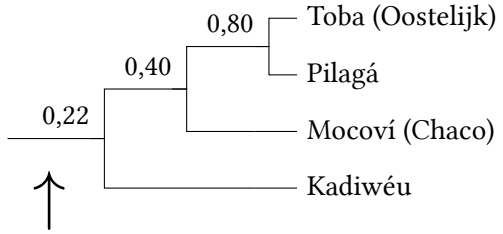
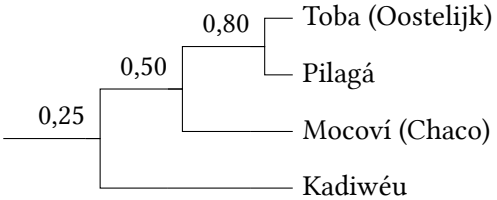
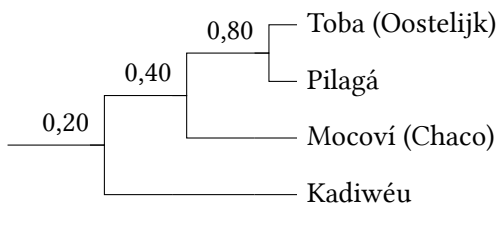
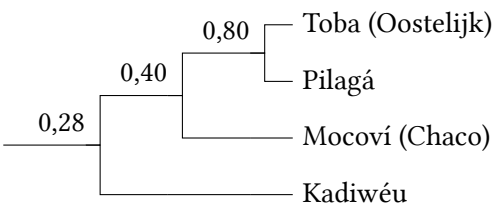
De gegevens zijn getranscribeerd in het Internationaal Fonetisch Alfabet. ˈ = hoofdklemtoon, ˌ = bijklemtoon (zwakker dan hoofdklemtoon), ◌ː = lange klank, ◌̆ = heel korte klank, X͡Y = X en Y worden

uitgesproken als een enkele klank, ◌́ = hoge toon, ◌̀ = lage toon, ◌̂ = dalende toon, ◌ˀ = voorgeglottaliseerde klank (voorafgegaan door een kortstondige blokkering van de luchtstroom in de keel), ◌’ = ejectieve klank (uitgesproken door kortstondig de luchtstroom in de keel te blokkeren), ◌̥ = stemloze klank, ◌̃ = genasaliseerde klank (uitgesproken door de neus), ◌̰ = krakende stem (een diep, krassend geluid), ◌ⁿ geeft aan dat er vóór de medeklinker wat lucht door de neus stroomt, ◌ʰ = geaspireerde medeklinker (uitgesproken met een zuchtje lucht), ◌ʷ = gelabialiseerde medeklinker (uitgesproken met afgeronde lippen), ◌ʲ = gepalatiseerde klank (uitgesproken met een deel van de tong bewogen dichtbij het harte verhemelte). **ɑ**, **æ**, **ɛ**, **ɪ**, **i̇**, **ɔ**, **ʊ**, **ʉ**, **ə**, **ʌ**, **ɒ**, **ɘ**, **ɣ**, **ɵ**, **ø** zijn klinkers. Andere speciale tekens zijn medeklinkers.

⚠ Kennis van de talen die in de opgave worden genoemd geeft geen voordeel bij het oplossen van de opgave.

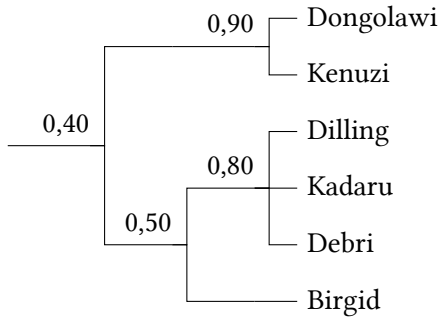
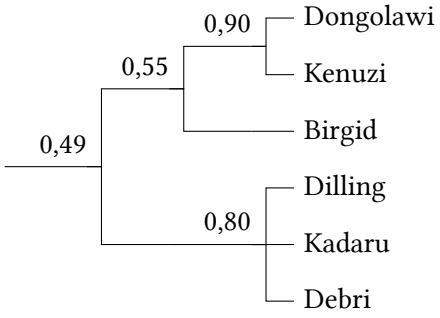
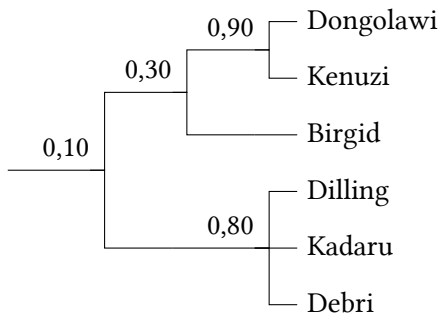
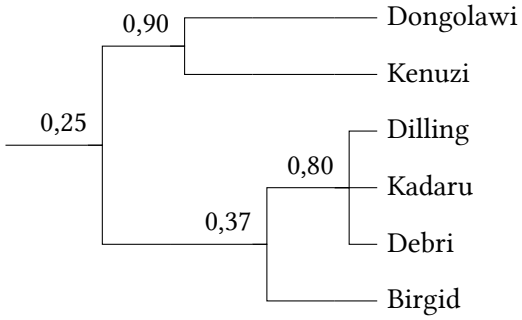
Deel I. Guaicurú-taalfamilie (Argentinië, Brazilië, Paraguay)

	Toba (Oostelijk)	Pilagá	Mocoví (Chaco)	Kadiwéu
wolk	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
vuur	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=ol:edi$_2$
vis	njaq$_1$	ˈnijaq$_1$	naʎin$_2$	nij:oɢo-d͡ʒegi$_3$
hoofd	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
doden	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
maan	ʔawoʁojk$_1$	ʔaˈwoʃojk$_1$	ʃiɾajɣo$_2$	ep:enaj$_3$
neus	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
zout	towe$_1$	olˈɣek$_2$	ʔwe$_1$	jok:i$_{-1}$
steen	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
tong	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʃat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

	Algoritme A	Algoritme B	
handmatig			Stabiliteitsindices: wolk 0,50 vuur 0,50 vis 0,50 hoofd 0,75 doden 1,00 maan 0,50 neus 1,00 zout 0,67 steen 0,75 tong 0,50
geautomatiseerd			Stabiliteitsindices: wolk 0,50 vuur 0,50 vis 0,75 hoofd 0,75 doden 1,00 maan 0,50 neus 1,00 zout 0,25 steen 0,75 tong 0,50

Deel II. Nubische taalfamilie (Egypte, Soedan)

	Dongolawi	Kenuzi	Dilling	Kadaru	Debri	Birgid
doden	ˈbɛː$_1$	beː$_1$	hur$_2$	wur-i$_2$	wur-i$_2$	filaːl-e$_1$
maan	uˈn-at-t$_1$	an-at-ti$_1$	nɔn-ti$_1$	nɔn-tu$_1$	nɔn-to$_1$	maːl$_2$
water	ˈɛss$_1$	essi$_1$	ɔti$_1$	ɔto$_1$	ɔtu$_1$	eɟi$_1$
geven	ˈtɪr$_1$	tir$_1$	ti$_1$	ti$_1$	ti$_1$	teː-n$_1$
goed	ˈsɛrɛː$_1$	sereː$_1$	ken$_2$	kɛn$_2$	kɛŋ$_2$	azze-n$_3$
wind	ˈturug$_1$	turug$_1$	irʃ-i$_2$	irʃ-o$_2$	irʃ-o$_2$	kurr-i$_3$
haar	ˈdɪl-tɪ$_1$	siːr$_2$	tel-ti$_1$	til-tu$_1$	til-tu$_1$	ur=dill-e$_1$
buik	ˈtuː$_1$	tuː$_1$	te-te$_2$	to$_1$	to$_1$	tuː$_1$
slapen	ˈnɛːr$_1$	neːr$_1$	ɟer$_1$	dwallɛli$_2$	ɟer-i$_1$	neːr-i$_1$
zon	ˈmasɪl$_1$	masɪl$_1$	ɛɟ$_2$	aɟu$_2$	eŋgal-to$_3$	ʔiːzi$_2$

	Algoritme A	Algoritme B	
handmatig			Stabiliteitsindices: doden 0,50 maan 0,83 water 1,00 geven 1,00 goed 0,50 wind 0,50 haar 0,83 buik 0,83 slapen 0,83 zon 0,50
geautomatiseerd			Stabiliteitsindices: doden 0,33 maan 0,50 water 0,50 geven 0,67 goed 0,50 wind 0,50 haar 0,83 buik 1,00 slapen 0,50 zon 0,50

(A) **(2 punten)** De medeklinker **ʁ** wordt uitgesproken als de Franse *r*, aan de achterkant van de tong. Tot welke klasse behoort deze medeklinker, en hoe hebben jullie dat bepaald?

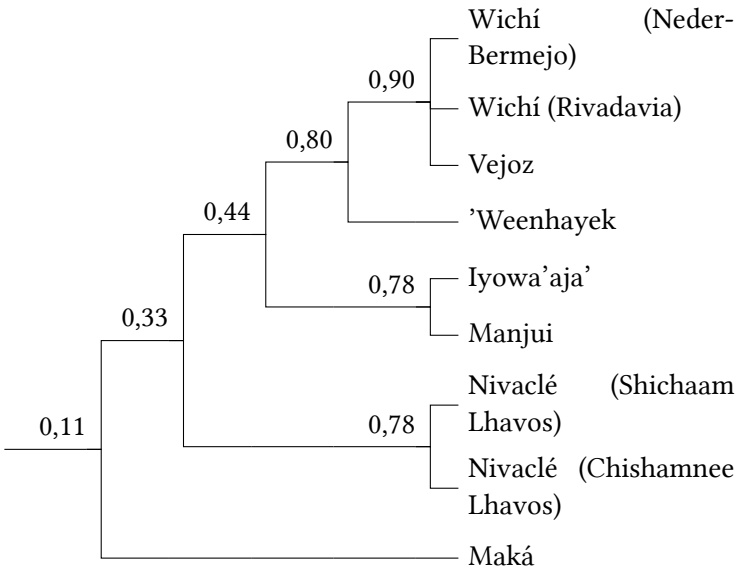
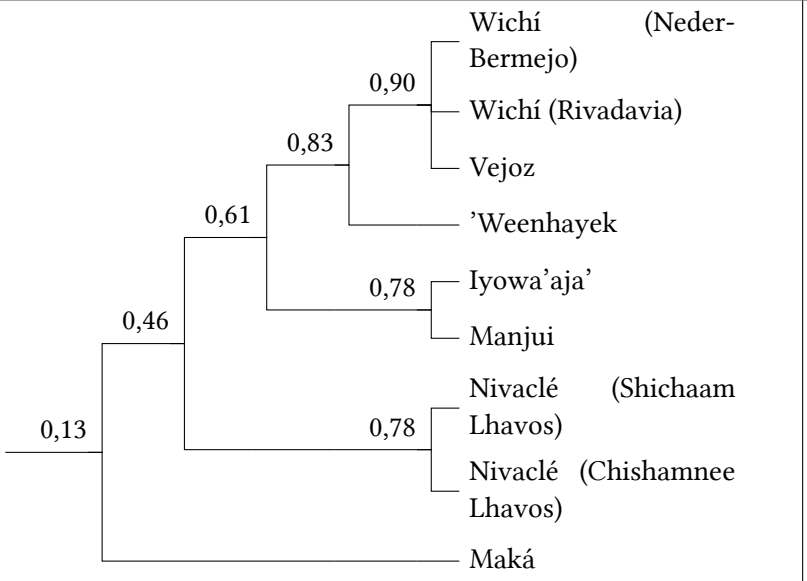
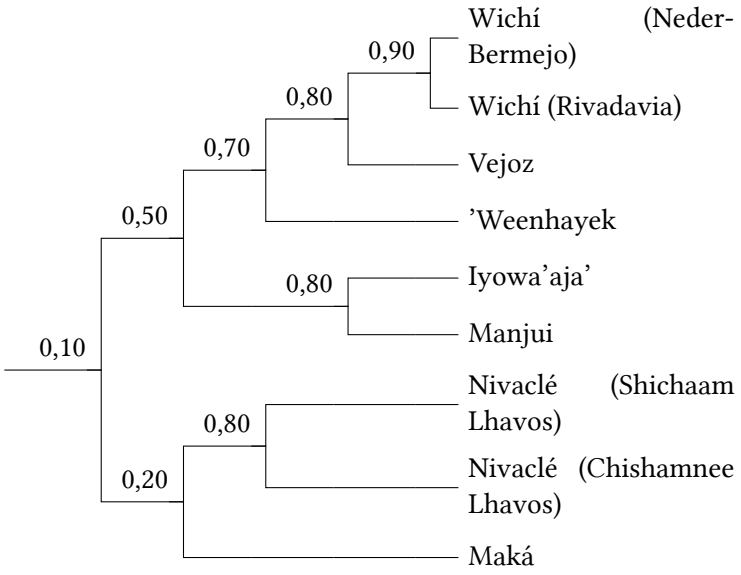
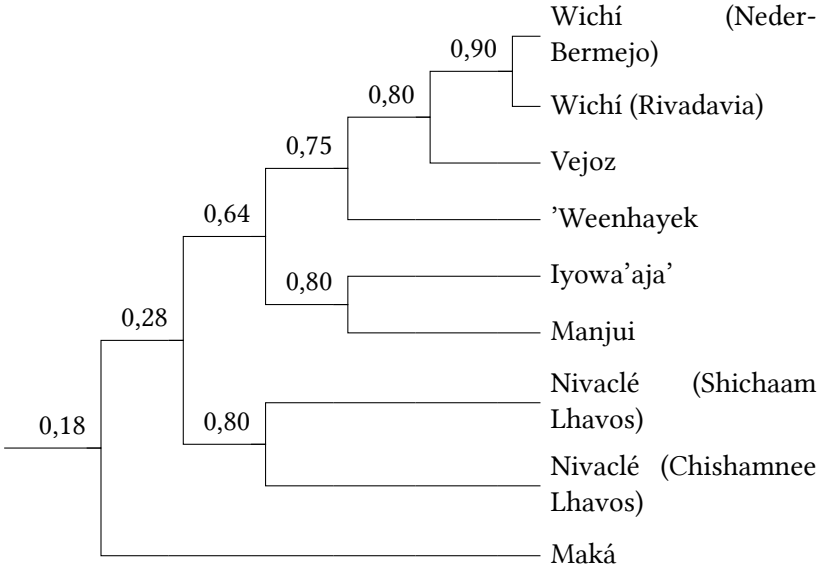
(B) **(2 punten)** De Nubische boom linksboven is slechts één van de twee mogelijke bomen voor deze combinatie van het algoritme en de annotatiemethode. Teken de andere mogelijke boom.

(C) **(2 punten)** De Nubische boom linksonder is slechts één van de twee mogelijke bomen voor deze combinatie van het algoritme en de annotatiemethode. Teken de andere mogelijke boom.

(D) **(2 punten)** De lexicostatistische afstand 0,49, die aan de wortel van de Nubische boom rechtsboven is toegewezen, is afgerond op twee decimalen (zoals enkele andere afstanden in deze opgave). Wat is de exacte afstand?

Deel III. Mataguayische taalfamilie (Argentinië, Bolivië, Paraguay)

	Wichí (Neder-Bermejo)	Wichí (Rivada-via)	Vejoz	’Weenhayek	Iyowa’aja’	Manjui	Nivaclé (Shichaam Lhavos)	Nivaclé (Chis-hamnee Lhavos)	Maká
vuur	ʔitoχ$_{1}$	ʔitɔx$_{1}$	ʔitɑh$_{1}$	ʔiːtɑχ$_{1}$	ˈh͡wat$_{2}$	ˈʔe͡itʲe$_{1}$	ʔitɑx$_{1}$	ʔitɑx$_{1}$	feʔt$_{2}$
vis	ʔwahat$_{1}$	wahat$_{1}$	wahat$_{1}$	ʔwaːhat$_{1}$	siˈʔjus$_{-1}$	ʃiˈʔjus$_{-1}$	saxet͡ʃ$_{-1}$	saxet͡ʃ$_{-1}$	sehet͡s$_{-1}$
voet	=pat͡ʃu$_{1}$	=qɔlɔ$_{2}$	=pat͡ʃo$_{1}$, =kɑla$_{2}$	=paːkʲʼoʔ$_{1}$	=ˈsat$_{3}$	=kaˈʔlaʔ$_{2}$	=ɸoʔ$_{4}$	=ɸoʔ$_{4}$	=fʼiʔ$_{5}$
water	ʔinot$_{1}$	ʔinɔt$_{1}$	wah$_{2}$	ʔinɑːt$_{1}$	ʔiˈn̪at$_{1}$	ʔaˈʔnat$_{1}$	jinɑʔt$_{1}$	jinɑʔt$_{1}$	iweliʔ$_{3}$
geven	=ʔweɲ̟-u$_{1}$	=weɲ̟-u$_{1}$	=ʔweɲ̟-o$_{1}$	=ʔweɲ̟-oʔ$_{1}$	=ˈwɛh͡n-aʔm$_{2}$	=ˈhajʔ$_{3}$, =ˈwɛn$_{2}$	=xut$_{4}$	=xut-ej$_{4}$	tis-ix$_{5}$
goed	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ʔis$_{1}$	ˈʔes$_{1}$	ˈʔe͡is$_{1}$	ʔis$_{1}$	ʔis$_{1}$	t=ejkʼun-ej$_{2}$
wind	ʔinwokʷ$_{1}$	ʔinwɔk$_{1}$	ʔihwokʷ$_{1}$	=jaːɬ$_{2}$, =xʷoxʷ$_{3}$	ˈh͡lah͡wuʔ$_{4}$	ˈh͡lah͡wuuʔ$_{4}$	ɬaβiʔm$_{5}$	ɬaβiʔm$_{5}$	tʼunikʼi$_{6}$
boom	haʔlo$_{1}$	halɔ$_{1}$	haʔlɑ$_{1}$	haʔlɑʔ$_{1}$	ʔaˈʔlaʔ$_{1}$	ʔaˈʔla-k$_{1}$	ʔaʔkxi-juk$_{2}$	jiʔk͡lɑʔ$_{1}$	naxka-k$_{3}$
haar	=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$_{5}$
doden	=lon$_{1}$	=lɔn$_{1}$	=lɑn$_{1}$	=lɑːn̪$_{1}$	=ˈlaʔan$_{1}$	=ˈlan$_{1}$	=k͡lɑn$_{1}$	=k͡lɑn$_{1}$	=lan$_{1}$

	Algoritme A	Algoritme B	
handmatig			Stabiliteitsindices: vuur 0,78 vis 1,00 voet 0,33 water 0,78 geven 0,44 goed 0,89 wind 0,33 boom 0,78 haar 0,67 doden 1,00
geautomatiseerd			Stabiliteitsindices: vuur 0,78 vis 0,44 voet 0,33 water 0,56 geven 0,67 goed 0,89 wind 0,22 boom 0,67 haar 0,67 doden 1,00

Deel IV. Mongoolse taalfamilie (Volksrepubliek China, Mongolië, Rusland)

(E) (10 punten) Bekijk de volgende woordenlijst. Bereken de stabiliteitsindices die met zowel de handmatige als de geautomatiseerde annotaties overeenkomen.

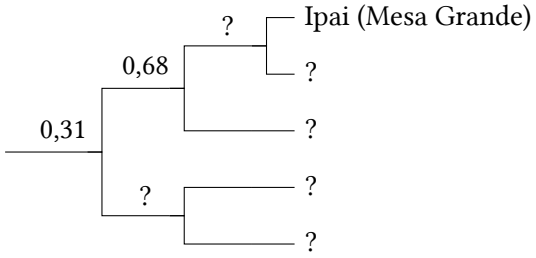
Om jullie te helpen zijn beide stabiliteitsindices voor het woord ‘alle’ al berekend. Deze indices zijn, in willekeurige volgorde, 0,36 en 0,40.

	Dagoer (Hailar)	Chamnigaans (Mantsjoe)	Boerjatisch (Khori)	Nieuw-Bargoetisch	Öölods	Khoshut	Kalmuks	Chalcha	Oerdoes	Oost-Joegoerisch	Bonan
alle	hɔː$_{1}$	bʊlt$_{2}$	bʊxiː$_{3}$	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$_{-2}$
schors	hai̯l̥s$_{1}$	qalihʊn$_{1}$	χɔltəhɔn$_{2}$	xaʲhuː$_{1}$	xolts$_{2}$	xalis$_{1}$	dursn$_{3}$	xɔɮt^{h}ɔ̆s$_{2}$	turusu$_{3}$	χalsən$_{1}$	arasun$_{4}$
buik	keːli$_{1}$	gətəhʉn$_{2}$	gedehen$_{2}$	gedyː$_{2}$	geːs$_{2}$	gets$_{2}$	gesn$_{2}$	gitĭs$_{2}$, xiwɮĭj$_{-1}$	ketysy$_{2}$	ketesən$_{2}$	kele$_{1}$
vogel	dəgi$_{-1}$	ɕiwan$_{1}$	ʃubuːn$_{1}$	ʃuwuː$_{1}$	ʃuvuː$_{1}$	ʃuwuː$_{1}$	ʃowun$_{1}$	ʃuwu$_{1}$	ʃuβuː$_{1}$	ʃuːn$_{1}$, pelt͡ʃɔr$_{2}$	bend͡ʑer$_{2}$
vuur	ɢali$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	ɢal$_{1}$	qal$_{1}$	qal$_{1}$	χal$_{1}$
weg	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$_{4}$
zout	hataː$_{1}$	dawhʊn$_{2}$	dabhan$_{2}$	dawuhuː$_{2}$	daws$_{2}$	daws$_{2}$	dawsn$_{2}$	tawsă$_{2}$	taβusu$_{2}$	taːpsən$_{2}$	dabsuŋ$_{2}$
zwemmen	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-$_{1}$
water	ɔsɔ$_{1}$	ʊxʊn$_{1}$	uhan$_{1}$	uːha$_{1}$	usn$_{1}$	us$_{1}$	usn$_{1}$	ʊsŏ$_{1}$	usun$_{1}$	q^{h}usun$_{1}$	sə$_{1}$
wind	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$_{1}$

Deel V. Yuman-taalfamilie (Mexico, VS)

(F) **(8 punten)** Bekijk de volgende woordenlijst. Hieronder volgt een boom geconstrueerd op basis van dezelfde lijst. Sommige gegevens (de namen van de talen en de lexicostatistische afstanden) ontbreken. Vul de lege plekken in. Geef aan of de boom handmatig of geautomatiseerd is, en ook of deze is gegenereerd met Algoritme A of B.

	Mojave	Cocopa	Yavapai	Tipai (Jamul)	Ipai (Mesa Grande)
kort	wena=wen-a_1	ˈxʌ=ʔut$_2$	ˈt͡ʃkr=ot-i_2	lə=ˈʔuɲ$_1$	mə=put-k_3
vogel	ʔit͡ʃ=i=jer$_1$	ˈʃa$_2$	ˈʔ=t͡ʃ=sa$_2$	aʔ=ˈʃa$_2$	ʔaː=ʂaː$_2$
bot	ɲ=a=s=ak$_1$	ˈɲ=j=aːk$_1$	ˈt͡ʃ=j=aːk-a_1	ˈak$_1$	aq$_1$
droog	i=roː-v-k$_1$	ˈʂ=ʔar$_2$	ˈru-β-i$_1$	ˈs=ʔaːj$_3$	saːj$_3$
vlees	kʷiːkʷay$_1$	ʔi=ˈmaːt͡ʃ$_2$	ˌkʷeː=ˈθo-β-a_3	ˈkʷak$_4$	kukʷaːj-p_1
nek	maʌaqe$_1$	ˈm=puk$_2$	ˈmlqí$_1$	iː=ˈpuk$_2$	iː=puk$_2$
zien	i=juː-k$_1$	ˈwiː$_2$	ˈʔuː$_1$	ˈwiːw$_2$	ə=wuːw$_2$
staart	iː=ʔar$_1$	ˈʃ=juʌ̊$_2$	ˈβ=hé$_3$	ʃə=ˈjuɬ$_2$	xə=juɬ$_2$
twee	havik-k$_1$	ˈx=wak$_1$	ˈh͡wâk-i$_1$	xə=ˈwak$_1$	xə=wak$_1$
jaar	huːðe$_1$	ˌmat-ˈkaːm$_2$	ˈʔ=t͡ʃʰur-a_3	mat-ˈwam$_2$	ʔaːnʲ$_{-1}$



(G) **(20 punten)** Enkele andere bomen werden gegenereerd voor de Yuman-taalfamilie, met de volgende lexicostatistische afstanden aan de wortel van elke boom (de lexicostatistische afstanden helemaal links van elke boom):

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

Teken elk van deze bomen. Geef voor elk van de bomen aan of deze handmatig of geautomatiseerd is, en ook of deze is gegenereerd met Algoritme A of B.

(H) **(3 punten)** Twee afstanden gegeven in opdracht **(G)** zijn afgerond op twee decimalen: 0,23 is afgerond vanaf 0,225. Welke andere afstand is afgerond, en wat was de precieze waarde ervan?

(I) **(4 punten)** Verklaar hoe de stabiliteitsindices worden berekend.

(J) **(5 punten)** Verklaar hoe de lexicostatistische afstanden worden berekend.

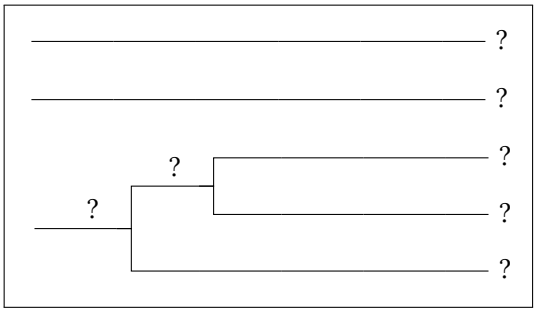
(K) **(4 punten)** Verklaar het verschil tussen Algoritmen A en B.

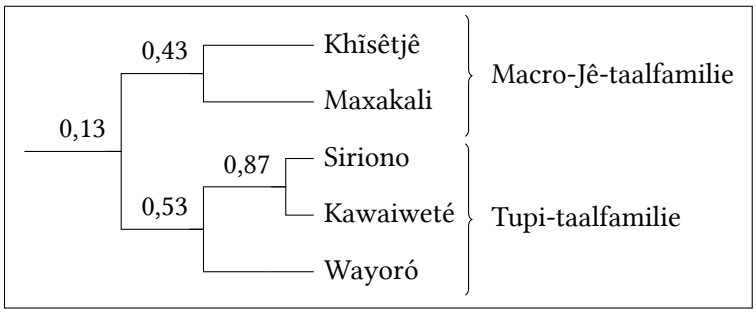
Deel VI. Macro-Jê-taalfamilie en Tupi-taalfamilie (Brazilië, Bolivië)

(L) (28 punten) De Macro-Jê talen en Tupi talen zijn twee grote taalfamilies van Zuid-Amerika. Enkele taalkundigen veronderstellen dat ze in de verte verwant zijn. Bekijk de volgende woordenlijsten.

	A	B	Γ	Δ	E
schors	e=ˈe-ke	hʷĩ=ˈkʰʌ	kʉp=ˈpɛ	mĩβ͡m=ˈt͡ɕaj	=ˈpɛ
buik	ˈe=ɾje	=ˈɻʰigi	=ãˈʉ̃n	=ˈtæj	=ɾɛˈwɛk
bloed	e=ˈɾuki	=kaˈⁿbɻo	=d͡ʒ=aˈʉ	=ˈhɛβ͡p	=ɾuˈɨ
branden	=ˈɾai̯	=ɾɔ=ˈkʰɹə̃	=poˈkʷa	mũ=...=ˈhaβ͡p	=ɾaˈpɨ
vet	e=ˈkiɾa	=ˈɻʰwəmɨ	=ˈd͡ʒ=ap	=ˈtuβ͡p	=ˈkap
voet	ˈe=i	=ˈhʷaji	=ˈβi	=pɒˈta	=ˈpɨ
hand	ˈe=o	=ɲɨˈkʰɹa	=ˈβo	=ˈɲĩβ͡m	=ˈpɔ
zwaar	e=ˈusi	=wɨˈɻʰĩ	=poˈti	=β͡pˈtəj	=pɔˈɨj
lever	ˈe=ja	=ˈⁿba	=piˈa	=t͡ɕɨβ͡pkɨˈnãj	=pɨˈʔa
nieuw	e=ˈjasu	=ˈⁿdɨwɨ	=paˈgop	=ˈtiβ͡p	=pɨaˈu
wortel	e=ˈɾao	=jaˈɻe	kʉp=kʉjoˈpɛ	mĩβ͡m=ɲĩβ͡m=t͡ɕaˈti̯ə	=ɾaˈpɔ
huid	ˈe=ɨ	=ˈkʰʌ	=ˈpɛ	=ˈt͡ɕaj	=ˈpit
staart	e=ˈɾoko͡i̯	=ˈⁿbɨ	=d͡ʒ=oˈkʷaj	=nãː=ˈkɨβ͡p	=ˈɾaj
wit	ˈe=ʃɨ̃	=jaˈkʰa	=d͡ʒiˈɾa	=β͡pˈdouɻ	=ˈsĩŋ
vleugel	e=ˈheo	=jaˈɻa	=pɛˈo	=ɲĩˈmãuɻ	=pɛˈpɔ, =jɨˈwa

Hieronder volgen twee bomen geconstrueerd op basis van dezelfde lijsten. Sommige gegevens (de namen van de talen en de lexicostatistische afstanden) ontbreken. Vul de lege plekken in. Geef voor elk van de bomen aan of deze handmatig of geautomatiseerd is, en ook of deze is gegenereerd met Algoritme A of B.





A	B	Γ	Δ	E
?	?	?	?	?

⚠ In deze opdracht zijn de handmatige annotaties en stabiliteitsindices opzettelijk weggelaten.

(M) (10 punten) Geautomatiseerde procedures op basis van Dolgopolsky-klassen kunnen onjuiste resultaten opleveren. In dit voorbeeld detecteert de geautomatiseerde procedure meer overeenkomsten tussen het Siriono en een bepaalde Macro-Jê taal (Khĩsêtjê) dan tussen het Siriono en andere Tupi talen. Stel een gewijzigde geautomatiseerde procedure voor die een juiste classificatie zou opleveren, als deze op de Macro-Jê en Tupi woordenlijsten boven zou worden toegepast, en beschrijf deze procedure *kort*.

⚠ Deze opdracht krijgt alleen een score als de beste teams gelijk staan.

De auteurs danken Alejandra Vidal, Maria Konoshenko, Ilya Gruntov en Jamthô Suyá voor het beantwoorden van hun vragen over specifieke talen. —*Andrey Nikulin, Milena Veneva*

Redactie: Ivan Derzhanski (technisch redacteur), Hugh Dobbs, Stanislav Gurevich, Boris Iomdin, Liam McKnight, Andrey Nikulin (hoofdredacteur), Aleksejs Peguševs, Jan Petr, Alexander Piperski, Maria Rubinstein, Milena Veneva, Elysia Warner.

Nederlandse tekst: Elysia Warner.

Succes!