21ST INTERNATIONAL LINGUISTICS OLYMPIAD

Team problem – Lexicostatistics

Andrey Nikulin and Milena Veneva

July 2024

Manual procedure

	Toba	Pilagá	Mocoví	Kadiwéu
cloud	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
fire	nodek$_1$	ˈd=oleʔ$_2$	norek$_1$	n=ol:edi$_2$
fish	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nij:oɢo-d͡ʒegi$_3$
head	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
to kill	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
moon	ʔawoʁojk$_1$	ʔaˈwoʕojk$_1$	ʃirajɣo$_2$	ep:enaj$_3$
nose	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
salt	towe$_1$	olˈɣek$_2$	ʔwe1	jok:i$_{-1}$
stone	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
tongue	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʕat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

	Toba	Pilagá	Mocoví	Kadiwéu	
cloud	1	1	2	3	2/4
fire	1	2	1	2	2/4
fish	1	1	2	3	2/4
head	1	1	1	2	3/4
to kill	1	1	1	1	4/4
moon	1	1	2	3	2/4
nose	1	1	1	1	4/4
salt	1	2	1	-1	2/3
stone	1	1	1	2	3/4
tongue	1	1	2	3	2/4

	Toba	Pilagá	Mocoví
Pilagá	8/10 = 0.80	–	–
Mocoví	6/10 = 0.60	4/10 = 0.40	–
Kadiwéu	2/9 = 0.22(2)	3/9 = 0.33(3)	2/9 = 0.22(2)

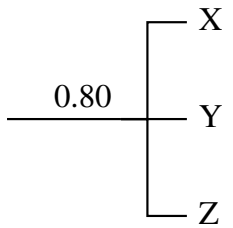
Borrowings (indicated by negative indices) are ignored in the manual procedure for all purposes.
Assignment I. Stability indices: maximum number of languages using cognate roots divided by total number of languages that have a native (non-borrowed) root.

Manual procedure – Algorithm A vs. Algorithm B

Assignment K. When clustering, Algorithm A uses the minimum value, Algorithm B uses the average.
Assignment J. Lexicostatistical distance: number of cognates divided by total number of comparable (non-borrowed) items. The maximum value during each iteration makes it to the tree, and the respective lects are grouped under a node. A new value is assigned to the node (see assignment K).

	Toba	Pilagá	Mocoví
Pilagá	8/10 = 0.80	–	–
Mocoví	6/10 = 0.60	4/10 = 0.40	–
Kadiwéu	2/9 = 0.22(2)	3/9 = 0.33(3)	2/9 = 0.22(2)

	X	Y	…
Y	0.80	–	–
Z	0.80	0.80	–
…	…	…	…

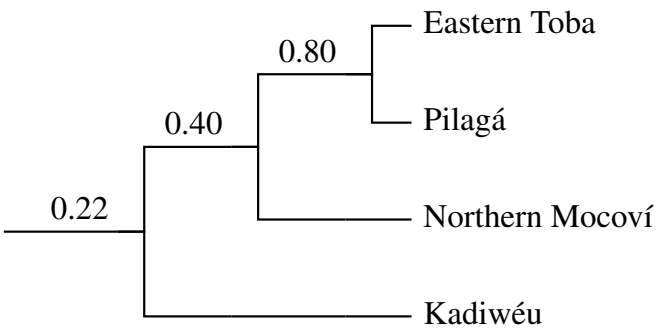


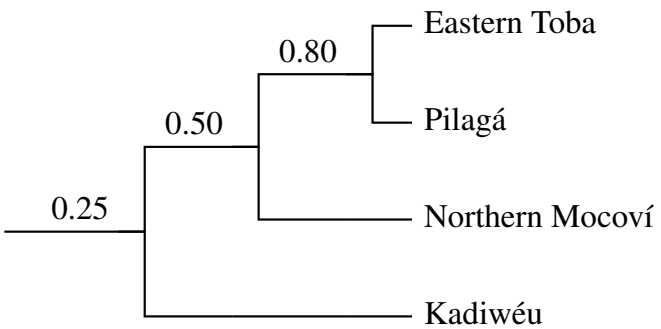
Toba + Pilagá = 0.80	Toba + Pilagá	Mocoví
Mocoví	min(0.60; 0.40) = 0.40	
Kadiwéu	min(0.22(2); 0.33(3)) = 0.22(2)	0.22(2)

Toba + Pilagá = 0.80	Toba + Pilagá	Mocoví
Mocoví	ave(0.60; 0.40) = 0.50	
Kadiwéu	ave(0.22(2); 0.33(3)) = 0.27(7)	0.22(2)

[Toba + Pilagá] + Mocoví = 0.40	[Toba + Pilagá] + Mocoví
Kadiwéu	min(0.22(2); 0.22(2)) = 0.22(2)

[Toba + Pilagá] + Mocoví = 0.50	[Toba + Pilagá] + Mocoví
Kadiwéu	ave(0.27(7); 0.22(2)) = 0.25





Automated procedure

	Toba	Pilagá	Mocoví	Kadiwéu
cloud	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
fire	nodek$_1$	ˈd=oleʔ$_2$	norek$_1$	n=ol:edi$_2$
fish	njaq$_1$	ˈnijaq$_1$	naʌin$_2$	nij:oɢo-d͡ʒegi$_3$
head	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
to kill	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
moon	ʔawoʁojk$_1$	ʔaˈwoʕojk$_1$	ʃirajɣo$_2$	ep:enaj$_3$
nose	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
salt	towe$_1$	olˈɣek$_2$	ʔwe1	jok:i$_{-1}$
stone	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
tongue	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʕat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

	Toba	Pilagá	Mocoví	Kadiwéu	
cloud	HK	HK	NK	RR	2/4
fire	NT	HR	NR	HR	2/4
fish	NK	NK	NR	NK	3/4
head	KK	KK	KK	HK	3/4
to kill	HR	HR	HR	HR	4/4
moon	HK	HK	SR	HP	2/4
nose	MK	MK	MK	MK	4/4
salt	TH	HR	HH	YK	1/4
stone	KH	KH	KH	WT	3/4
tongue	HS	HS	HR	HK	2/4

	Toba	Pilagá	Mocoví
Pilagá	8/10 = 0.80	–	–
Mocoví	4/10 = 0.40	4/10 = 0.40	–
Kadiwéu	3/10 = 0.30	4/10 = 0.40	2/10 = 0.20

1) naweɣelek → NK

2) ʔwe → HH

3) X͡Y → Y:

=a t͡ʃ-aʁat → H S

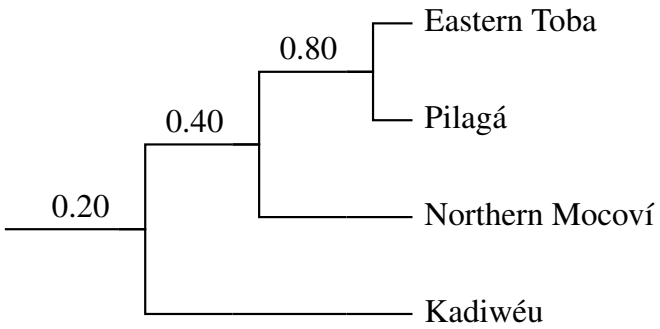
P	p b ɓ ɸ β f v	M	m ɱ
T	t d ɗ θ ðð ʈ ɖ	N	n ɳ ɲ ŋ
S	s z ʃ ʒ ʂ ʐ ɕ ʑ c ɟ	Q	t͡ɬ d͡ɮ
Y	j ç (root-initially)	R	r ɾ l ɬ ɮ ɭ ʎ ɬ
W	w ʍ (root-initially)	K	k g x ɣ q ɢ χ
H	ħ ʕ ʜ ʢ ʡ h ɦ ʔ, vowels, and j ç w ʍ (except root-initially)		

Automated procedure – Algorithm A vs. Algorithm B

	Toba	Pilagá	Mocoví
Pilagá	8/10 = 0.80	–	–
Mocoví	4/10 = 0.40	4/10 = 0.40	–
Kadiwéu	3/10 = 0.30	4/10 = 0.40	2/10 = 0.20

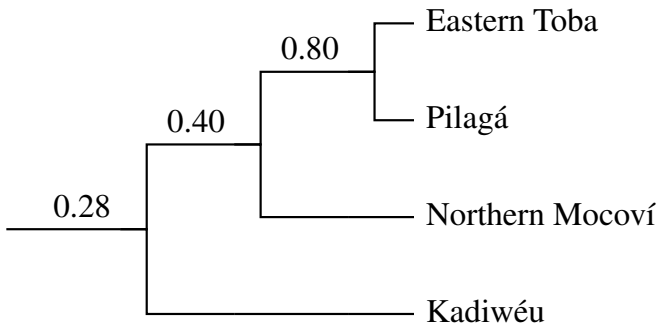
Toba + Pilagá = 0.80	Toba + Pilagá	Mocoví
Mocoví	min(0.40; 0.40) = 0.40	
Kadiwéu	min(0.30; 0.40) = 0.30	0.20

[Toba + Pilagá] + Mocoví = 0.40	[Toba + Pilagá] + Mocoví
Kadiwéu	min(0.30; 0.20) = 0.20



Toba + Pilagá = 0.80	Toba + Pilagá	Mocoví
Mocoví	ave(0.40; 0.40) = 0.40	
Kadiwéu	ave(0.30; 0.40) = 0.35	0.20

[Toba + Pilagá] + Mocoví = 0.40	[Toba + Pilagá] + Mocoví
Kadiwéu	ave(0.35; 0.20) = 0.275



Assignment A. Consonant class of ʁ.

	Toba	Pilagá	Mocoví	Kadiwéu
cloud	l=ʔok$_1$	ˈlo=ʔok$_1$	naweɣelek$_2$	lol:adi$_3$
fire	nodek$_1$	ˈd=oleʔ$_2$	noɾek$_1$	n=ol:edi$_2$
fish	njaq$_1$	ˈnijaq$_1$	naʎin$_2$	nij:oɢo-d͡ʒegi$_3$
head	=qajk$_1$	=ˈqajk$_1$	=qaik$_1$	=ak:ilo$_2$
to kill	=alawat$_1$	=aˈla:t$_1$	=alawat$_1$	=el:owadi$_1$
moon	ʔawoʁojk$_1$	ʔaˈwoʃojk$_1$	ʃɪrajɣo$_2$	ep:enaj$_3$
nose	=mik$_1$	=ˈmik$_1$	=mik$_1$	=m:iq:o$_1$
salt	towe$_1$	olˈɣek$_2$	ʔwe1	jok:i$_{-1}$
stone	qaʔ$_1$	ˈqaʔ$_1$	qaʔ$_1$	wet:iɢa$_2$
tongue	=at͡ʃ-aʁat$_1$	=aˈt͡ʃ-aʃat$_1$	=oʔleɣ-aʁan-aʁat$_2$	=ok:el:i$_3$

0.80 = 8/10

But...

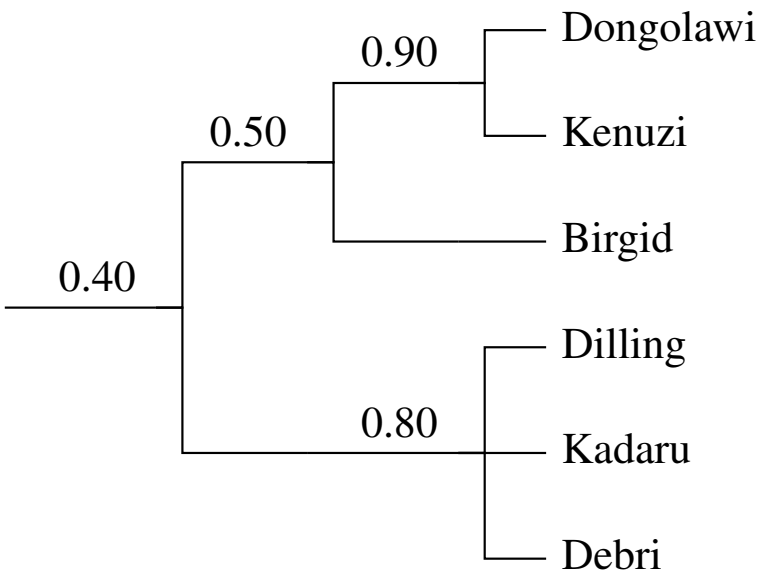
ʔaˈwoʃojk → HK

ʔawoʁojk (if ʁ ∉ H) / ʔawoʁojk (if ʁ ∈ H) → HK

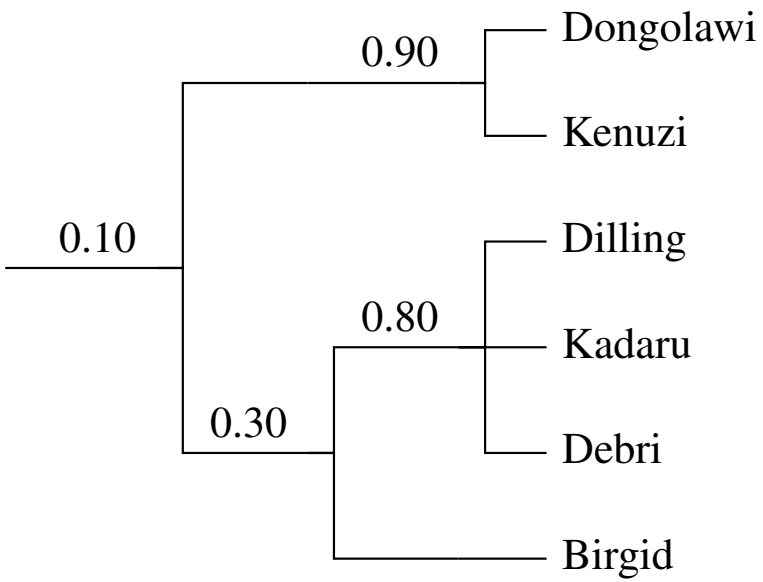
Then, *ʁ* belongs to consonant class **K** (or **H**, which is less likely, as *ʁ* is known to be articulated at the back of the tongue).

Assignments B, C. Other Algorithm A trees for Nubian.

Assignment B.
Manual tree, Algorithm A.



Assignment C.
Automated tree, Algorithm A.

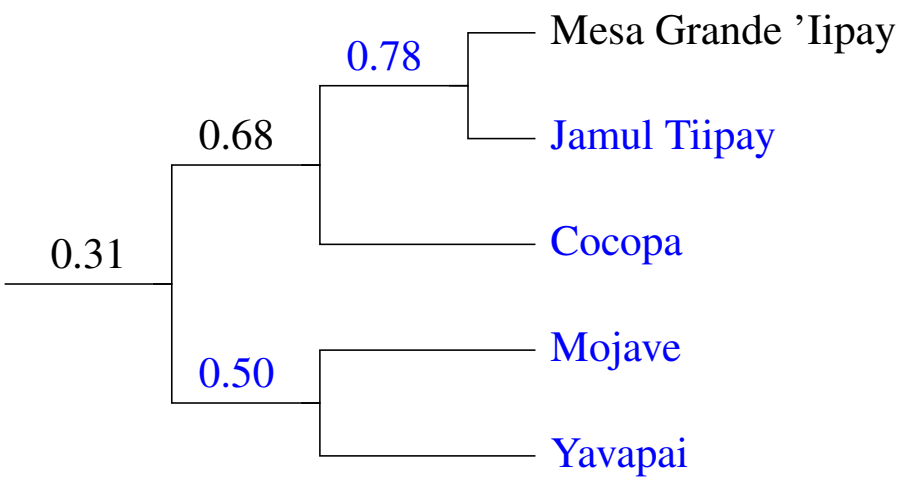


Assignment D. 0.491(6) = 59/120 = ave(ave(0.45; 0.45; 0.55); 0.50).

Assignments F, G.

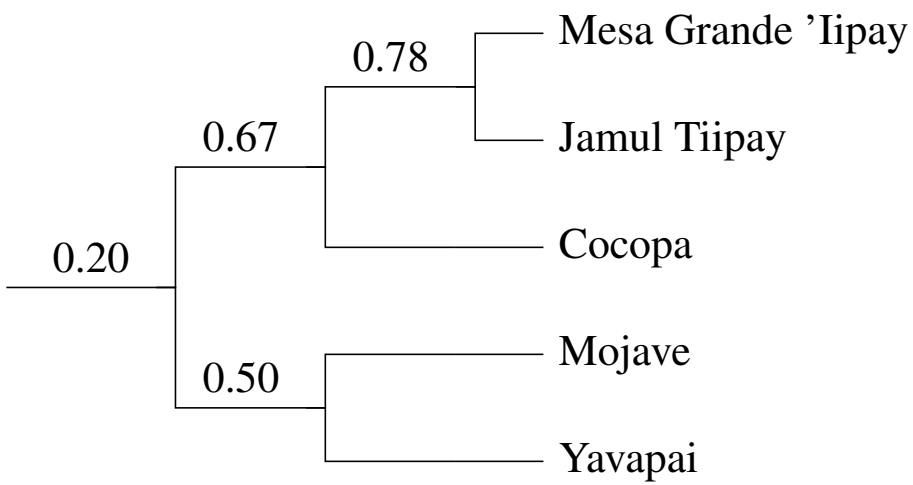
Assignment F.

Manual tree, Algorithm B.



Assignment G.

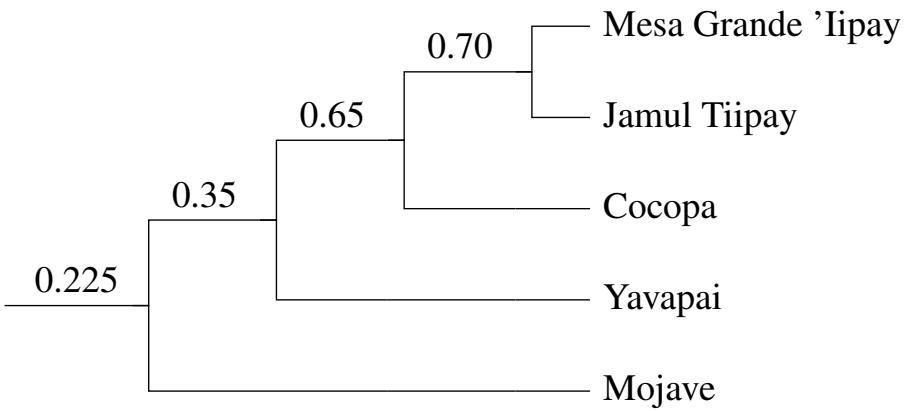
1. Distance = 0.20. Manual tree, Algorithm A.



Assignments G, H.

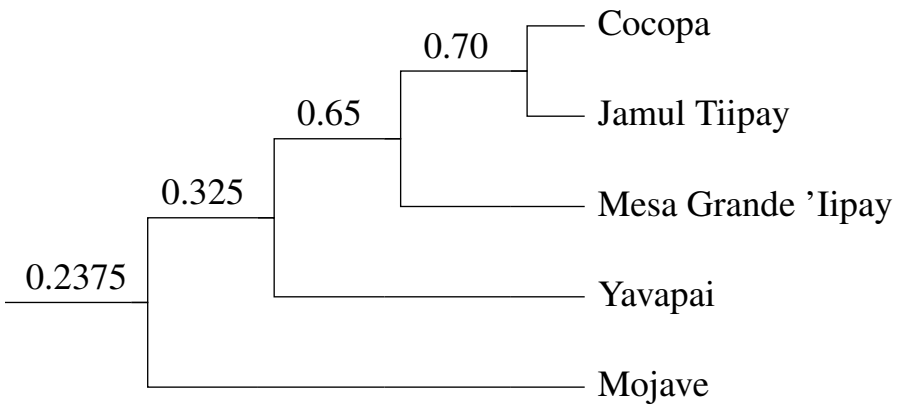
Assignment G.

2. Distance = 0.23. Automated tree, Algorithm B.



Assignments G, H.

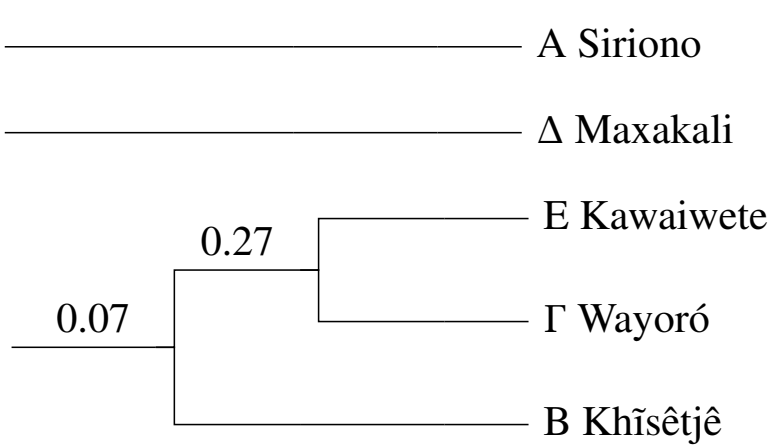
3. Distance = 0.24 **= 0.2375**. Automated tree, Algorithm B.



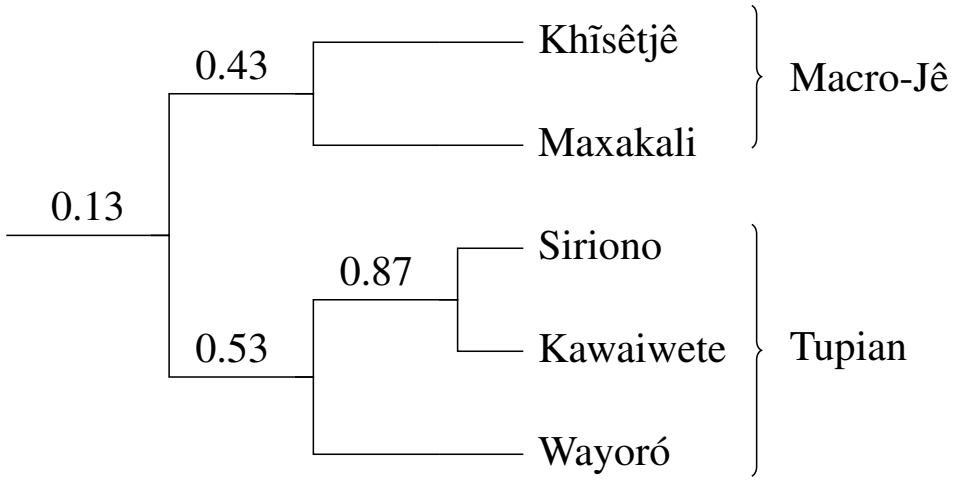
Assignment H. 0.2375 = ave(0.175; 0.30).

Assignment L.

Automated tree, Algorithm A.



Manual tree, Algorithm A.



A	B	Γ	Δ	E
Siriono	Khĩsêtjê	Wayoró	Maxakali	Kawaiwete

How did we check the correctness of the task?

```
Distance matrix for Nubian.txt
0.9
0.5  0.4
0.5  0.4  0.8
0.6  0.5  0.8  0.8
0.6  0.5  0.5  0.5  0.5
stability indeces [0.5, 0.8333333333333334, 1.0, 1.0, 0.5, 0.5, 0.8333333333333334, 0.8333333333333334, 0.8333333333333334, 0.5]
==============================
Manual trees creation:
==============================
Max value =  0.9
Number of max values =  1
Merge  [Dongolawi + Kenuzi]  with distance  0.9
==============================
Max value =  0.8
Number of max values =  3
Merge  [Kadaru + Debri + Dilling]  with distance  0.8
==============================
Max value =  0.55
Number of max values =  1
Merge  [[Dongolawi + Kenuzi] + Birgid]  with distance  0.55
==============================
Max value =  0.4916666666666667
Number of max values =  1
Merge  [[[Dongolawi + Kenuzi] + Birgid] + [Kadaru + Debri + Dilling]]  with distance  0.4916666666666667
```

Thank you for your attention!