

Essai de classification des langues d'après la description de M. Ruhlen

A. Guénoche
IML – CNRS
163 Av. de Luminy, 13009 Marseille
guenoche@iml.univ-mrs.fr

Nous avons essayé de classer les langues décrites par M. Ruhlen, en appliquant la méthode de classification par densité, présentée dans un texte à part.

1 Les données

Nous sommes partis de son tableau de données binaires qui caractérise 100 langues par 274 attributs binaires. Ces 100 langues sont réparties en 17 familles, plus une famille d'"isolées" que nous avons éliminée. Il reste alors 94 langues. Les familles considérées sont

AA = AFRO-ASIATIC
AL = ALTAIC
AM = AMERIND
AS = AUSTRIC
AU = AUSTRALIAN
CA = CAUCASIAN
DR = DRAVIDIAN
EA = ESKIMO-ALEUT
IE = INDO-EUROPEAN
IP = INDO-PACIFIC
KH = KHOISIAN
ND = NA-DENE
NK = NIGER-KORDOFANIAN
NS = NILO-SAHARIAN
PS = CHUKCHI-KAMCHATKAN
ST = SINO-TIBETAN
UR = URALIC

Les 274 attributs binaires sont répartis en 180 attributs phonologiques qui indiquent la présence-absence de phonèmes dans la langue et 94 attributs grammaticaux.

2 La méthode

La première opération consiste à calculer une valeur de distance entre toute paire de langues. La distance choisie est la distance de Jaccard, qui ne comptabilise pas les attributs deux fois absents, puisque "ne pas posséder un caractère" ne rapproche en rien les langues considérées. Cette distance est donc inversement proportionnelle au nombre d'attributs partagés rapporté au nombre d'attributs présents dans l'une ou l'autre langue.

La méthode de classification par densité a ensuite été appliquée et, pour un graphe de degré initial égal à 1, elle a produit au niveau des noyaux et des classes étendues 14 classes. Pour un graphe de degré 2, il ne reste plus que 9 classes et ce nombre va décroissant très vite. Dans la partition finale, nous avons rajouté une classe composée des éléments non affectés, car non connectés au reste du graphe. Les résultats figurent dans l'annexe I.

Nous avons mesuré les critères de qualité, définis pour valider la méthode de classification, en comparant les classes obtenues aux familles linguistiques données. Les valeurs des critères appliqués à nos classes comparées aux 17 familles sont successivement :

Noyaux :	Taux d'élément bien classés :	0.72
	Taux de paires réunies :	0.48
Classes :	Taux d'élément bien classés :	0.65
étendues	Taux de paires réunies :	0.32
Classes :	Taux d'élément bien classés :	0.62
complètes	Taux de paires réunies :	0.23

Ces résultats sont plutôt mauvais compte tenu de ce que l'on peut attendre d'une table binaire dans laquelle figure une information différenciée sur les familles (attributs ou combinaison d'attributs spécifiques).

Pour vérifier qu'il ne s'agissait pas d'un défaut de la méthode de classification, nous avons testé d'autres algorithmes de type allocation-recentrage qui cherchent à minimiser un critère, d'inertie ou de séparation, sur l'ensemble des partitions à nombre de classes fixé. En s'en tenant au nombre de classes proposé par la méthode de densité, nous n'avons pas obtenus de meilleurs résultats (partition plus proche des familles), voire des résultats nettement moins bons.

3 Explication des résultats

Reste à comprendre pourquoi les langues se regroupent dans des classes qui ne sont pas des familles linguistiques. (Il s'agit sans doute d'influences dues au voisinage géographique).

Pour ceci, nous avons réalisé un nouveau programme qui pour chaque langue et chaque famille établit deux choses :

- le pourcentage d'attributs qu'elle possède en commun avec les langues de sa famille et
- la liste des langues qui n'appartiennent pas à sa famille, mais avec laquelle elle possède un pourcentage supérieur au maximum observé dans sa famille.

Exemple : La famille des langues Afro-asiatiques (AA) ne se retrouve pas dans nos classes.

AA-Tuareg	0	31	36	44	50
AA-Tera	31	0	54	46	42
AA-Walamo	36	54	0	65	48
AA-Afar	44	46	65	0	51
AA-Iraqi	50	42	48	51	0

Langues plus proches que la plus proche voisine

AA-Tuareg	IE-Gaelic (52)	IE-Rumanian (55)
AA-Tera	IE-Russian (60)	IP-Andamanese (55) NK-Shona (56) NS-Songhai (59)
	NS-Fur (59)	NS-Berta (56)
AA-Walamo		
AA-Afar	AS-Banjalong (68)	
AA-Iraqi	IE-Albanian (55)	IE-Gaelic (52) IE-Rumanian (55)

Le Touareg possède 50% d'attributs communs avec l'Iraqi ; c'est la langue la plus proche de sa famille. Mais il ressemble plus à deux langues Indo-européennes, le Gaélique et le Roumain

avec qui il partage respectivement 52 et 55 pourcents d'attributs. Pour toutes les autres langues de la famille, il en est de même, sauf le Walamo qui possède 65% d'attributs commun avec l'Afar et aucune autre langue de l'ensemble ne fait mieux.

Les résultats sur les autres familles sont indiqués sans plus de commentaires dans l'annexe II.

Les familles Altaïc (AL) et Amérindienne (AM) sont particulièrement perturbées et dans des familles plus homogènes, certaines langues paraissent exotiques (Hindi ou l'Arménien dans les Indo-européennes, l'Elamite chez les dravidiennes, le Hua chez les Khoisian, etc).

Ces nombreuses proximités entre langues de familles différentes expliquent pourquoi une méthode basée sur des distances ne peut retrouver ces familles. Leur principe est de regrouper ce qui est voisin, et la ressemblance forte entre le Géorgien et l'Arménien, fait qu'on les retrouvera classées ensemble, même si elles appartiennent aux familles caucasiennes et indo-européenne.

Ce voisinage perturbateur entre langues hétérogènes est confirmé par une seconde étude, dans laquelle on a cherché à mesurer la moyenne des proximités de chaque langue avec les 17 familles. Il apparaît que pour 42 langues, elles sont plus proches – en moyenne - d'une autre famille que de la leur ! Ces résultats sont visibles dans l'Annexe III. Le Hua, déjà repéré, ressemble plus à n'importe quel autre groupe qu'au sien ! Les seules familles qui ne supportent pas d'exceptions sont les langues Caucasiennes (CA), Esquimo-Aléoutienne (EA), Na-Dene (ND), Chukchi-Kamchatka (PS) et Ouralienne (UR), toutes très peu représentées.

4 Conclusions

Selon ces données il est très difficile, voire impossible de retrouver les familles. Nous avons essayé de sélectionner soit les attributs phonologiques seuls, soit les autres ; les résultats ont été encore pires. La sélection qui donne une légère amélioration est celle qui ne conserve que les attributs qui sont majoritaires dans au moins une famille – et qui contribuent donc à l'homogénéité de celle-ci. Mais les résultats ne sont pas satisfaisants.

Il faudrait donc soit changer d'espace de représentation – définir d'autres attributs - soit chercher une caractérisation famille par famille – une contre toute les autres. On risque d'obtenir ainsi des règles très ad hoc.

Annexe I

Distance File : Ruh194

Nb. of edges 77 (percentage 0.018), Average Degree 1.6

Nb. of classes 14

Kernels (Nb. of elements 43)

Percentage of edges 0.580

Class 1

AA-Walamo AA-Afar AS-Banjalang

Class 2

AL-Khalkha AL-Evenki IP-Andamanese PS-Chukchi

Class 3

AL-Bashkir AM-Carib UR-Yukaghir

Class 4

AS-WikNunkan AS-Pitjantjatj

Class 5

AU-Nancowry AU-Temiar AU-Mor2 AU-Arosi IP-Reefs

Class 6

AU-Atayal AU-WBukidnon AU-Easter

Class 7

CA-Georgian CA-Kabardian IE-Armenian

Class 8

EA-Aleut EA-Yupik

Class 9

IE-Albanian IE-Rumanian

Class 10

IE-Russian IE-Gaelic

Class 11

IP-Waskia IP-Woisika ST-Khami ST-Lahu

Class 12

NK-Moro NK-Fula NK-Shona

Class 13

NK-Loma NK-More NS-Songhai NS-Berta

Class 14

NS-Fur NS-Kanuri ST-Bodo

Extended classes (Nb. of elements 74)

Percentage of edges 0.328

Class 1

AA-Walamo AA-Afar AL-Japanese AS-Banjalang IP-Kwoma

IP-Savosavo

Class 2

AL-Khalkha AL-Evenki AS-Tiwi IP-Andamanese PS-Chukchi

PS-Kamchadal

Class 3

AL-Bashkir AL-Korean AM-Carib UR-Yukaghir UR-Yenets

UR-Finnish

Class 4

AS-Maranungu AS-WikNunkan AS-Pitjantjatj

Class 5

AM-Tuscarora AU-Cambodian AU-Nancowry AU-Temiar AU-Thai

AU-Mor2

Class 6

AM-Jacaletc AU-Atayal AU-WBukidnon AU-Easter

Class 7

CA-Georgian CA-Kabardian CA-Xnalug IE-Armenian

Class 8

EA-Aleut	EA-Yupik			
Class 9				
AA-Tuareg	AA-Iraqi	AM-Itonama	IE-Albanian	IE-Rumanian
Class 10				
AA-Tera	IE-Russian	IE-Gaelic	IE-English	
Class 11				
AM-Mojave	AU-Miao	IP-Kate	IP-Kunimaipa	IP-Waskia
IP-Woisika	ST-Tibetan	ST-Khami	ST-Lahu	
Class 12				
NK-Moro	NK-Fula	NK-Awing	NK-Shona	
Class 13				
KH-Kung	KH-Hua	NK-Loma	NK-More	NS-Songhai
NS-Berta	NS-Bagirma			
Class 14				
AM-Tucuna	AU-Mamay	DR-Elamite	NS-Fur	NS-Kanuri
ST-Bodo				

Final classes (Nb. of elements 94)
Percentage of edges 0.287

Class 1				
AA-Walamo	AA-Afar	AL-Japanese	AS-Banjalong	AU-Khasi
IP-Kwoma	IP-Savosavo			
Class 2				
AL-Khalkha	AL-Evenki	AS-Tiwi	IP-Andamanese	PS-Chukchi
PS-Kamchadal	ST-Mandarin			
Class 3				
AL-Bashkir	AL-Korean	AM-Carib	UR-Yukaghir	UR-Yenets
UR-Finnish				
Class 4				
AS-Maranungu	AS-WikNunkan	AS-Pitjantjatj		
Class 5				
AM-Tuscarora	AU-Cambodian	AU-Nancowry	AU-Temiar	AU-Thai
AU-Mor2	AU-Arosi	IP-Reefs	ST-Sgaw	
Class 6				
AM-Jacaletc	AU-Atayal	AU-WBukidnon	AU-Easter	
Class 7				
AM-Halkomelen	CA-Georgian	CA-Kabardian	CA-Xnalug	IE-Armenian
Class 8				
EA-Aleut	EA-Yupik			
Class 9				
AA-Tuareg	AA-Iraqi	AM-Itonama	IE-Albanian	IE-Rumanian
Class 10				
AA-Tera	IE-Russian	IE-Gaelic	IE-English	
Class 11				
AM-Mojave	AU-Miao	IP-Kate	IP-Kunimaipa	IP-Waskia
IP-Woisika	ST-Tibetan	ST-Khami	ST-Lahu	
Class 12				
NK-Moro	NK-Fula	NK-Awing	NK-Shona	
Class 13				
KH-Kung	KH-Hua	NK-Loma	NK-More	NS-Songhai
NS-Berta	NS-Bagirma			
Class 14				
AM-Mixtec	AM-Tucuna	AU-Bru	AU-Mamay	DR-Elamite
NK-Banda	NS-Fur	NS-Kanuri	NS-Anuak	ST-Bodo
Class 15				
AS-Djingili	AU-Yao	AU-Kharia	AU-Lakkia	DR-Brahui
DR-Malto	DR-Tamil	IE-Hindi	KH-Korana	ND-Tlingit
ND-Carrier	ND-Navajo			

Annexe II

Pour chaque famille de langues sont établis :

- Le pourcentage d'attributs communs entre chaque langue
- Pour chaque langue, la liste des langues en dehors de sa famille qui possèdent avec elle un pourcentage de similitude supérieur au maximum observé dans sa famille.

Famille 1, nb. d'elements 5

AA-Tuareg	0	31	36	44	50
AA-Tera	31	0	54	46	42
AA-Walamo	36	54	0	65	48
AA-Afar	44	46	65	0	51
AA-Iraqi	50	42	48	51	0

Langues plus proches que le plus proche voisin

AA-Tuareg	IE-Gaelic (52) IE-Rumanian (55)
AA-Tera	IE-Russian (60) IP-Andamanese (55) NK-Shona (56) NS-Songhai (59) NS-Fur (59) NS-Berta (56)
AA-Walamo	
AA-Afar	AS-Banjalang (68)
AA-Iraqi	IE-Albanian (55) IE-Gaelic (52) IE-Rumanian (55)

Famille 2, nb. d'elements 5

AL-Bashkir	0	54	54	50	52
AL-Khalkha	54	0	74	39	45
AL-Evenki	54	74	0	43	48
AL-Korean	50	39	43	0	44
AL-Japanese	52	45	48	44	0

Langues plus proches que le plus proche voisin

AL-Bashkir	AM-Carib (59) EA-Aleut (57) IP-Andamanese (63) IP-Waskia (56) NS-Fur (56) PS-Chukchi (55) ST-Bodo (64) UR-Yukaghir (66) UR-Yenets (56)
AL-Khalkha	
AL-Evenki	
AL-Korean	AM-Carib (52) CA-Georgian (52) IP-Andamanese (53) PS-Chukchi (54) PS-Kamchadal (51) ST-Khami (52) UR-Yukaghir (56)
AL-Japanese	AA-Walamo (60) AA-Afar (59) AM-Carib (55) DR-Tamil (57) IP-Andamanese (53) IP-Kwoma (55) UR-Yukaghir (53)

Famille 3, nb. d'elements 8

AM-Halkomelen	0	32	26	37	32	31	25	20
AM-Tuscarora	32	0	34	31	31	38	39	39
AM-Mojave	26	34	0	32	40	31	46	40
AM-Jacaletc	37	31	32	0	35	38	39	29
AM-Mixtec	32	31	40	35	0	48	37	48
AM-Itonama	31	38	31	38	48	0	42	42
AM-Carib	25	39	46	39	37	42	0	52
AM-Tucuna	20	39	40	29	48	42	52	0

Langues plus proches que le plus proche voisin

AM-Halkomelen	CA-Xnalug (39) IE-Rumanian (39)
AM-Tuscarora	AU-Mor2 (43) DR-Malto (42)
AM-Mojave	AS-WikNunkan (51) AU-Cambodian (50) AU-Atayal (48) AU-Mor2 (48)

AU-Easter (49) IP-Andamanese (47) IP-Kate (47) IP-Waskia (60)
 IP-Woisika (52) KH-Hua (50) ND-Carrier (50) NK-Fula (47)
 NK-More (55) NS-Songhai (54) NS-Fur (57) NS-Kanuri (56)
 NS-Berta (49) PS-Chukchi (53) ST-Sgaw (53) ST-Tibetan (47)
 ST-Khami (60) ST-Bodo (50) UR-Yukaghir (55)
 AM-Jacaletc AA-Walamo (42) AL-Korean (43) AU-Bru (43) AU-Atayal (50)
 AU-Mamay (48) AU-Mor2 (43) AU-Arosi (41) AU-Easter (49)
 CA-Georgian (49) CA-Kabardian (40)
 IE-Armenian (45) ND-Carrier (44) NK-Fula (45) NS-Berta (43)
 PS-Kamchadal (40) ST-Sgaw (41)
 AM-Mixtec AU-Temiar (49) AU-Mamay (55) AU-Arosi (51) IE-Gaelic (50)
 IE-Rumanian (51) IP-Reefs (50) NK-Moro (50) NK-Fula (52)
 NS-Anuak (49)
 AM-Itonama AU-Mamay (49) IE-Gaelic (49)
 IE-Rumanian (54) AM-Carib AA-Walamo (56) AA-Afar (60)
 AL-Bashkir (59) AL-Khalkha (63) AL-Evenki (64) AL-Japanese (55)
 AS-Banjalang (54) AU-Kharia (54) AU-WBukidnon (57) AU-Mamay
 (58) AU-Arosi (56) AU-Easter (54) DR-Malto (57) IP-Andamanese
 (64) IP-Kate (54) IP-Kunimaipa (53) IP-Waskia (60) IP-Woisika
 (64) IP-Savosavo (55) NK-Loma (56) NS-Fur (62) NS-Kanuri (53)
 PS-Chukchi (59) ST-Bodo (65) UR-Yukaghir (71) UR-Yenets (60)
 UR-Finnish (53)
 AM-Tucuna NS-Fur (55)

Famille 4, nb. d'elements 6

AS-Tiwi	0 26 39 38 43 41
AS-Djingili	26 0 31 29 30 34
AS-Maranungu	39 31 0 64 41 57
AS-WikNunkan	38 29 64 0 35 77
AS-Banjalang	43 30 41 35 0 32
AS-Pitjantjatj	41 34 57 77 32 0

Langues plus proches que le plus proche voisin

AS-Tiwi AM-Carib (46) AU-WBukidnon (46) EA-Yupik (46) IP-Andamanese
 (44) IP-Savosavo (48) PS-Chukchi (50) ST-Bodo (45)
 AS-Djingili KH-Korana (43)
 AS-Maranungu
 AS-WikNunkan
 AS-Banjalang AA-Walamo (50) AA-Afar (68) AL-Bashkir (48) AL-Khalkha (49)
 AL-Evenki (49) AL-Japanese (47) AM-Carib (54) DR-Malto (52)
 DR-Tamil (46) IE-Russian (49) IE-Gaelic (47) IP-Andamanese (61)
 IP-Waskia (50) IP-Kwoma (45) IP-Savosavo (58) NK-More (49)
 NS-Songhai (48) NS-Fur (53) PS-Chukchi (50) ST-Bodo (59)
 UR-Yukaghir (58) UR-Yenets (46)
 AS-Pitjantjatj

Famille 5, nb. d'elements 16

AU-Miao	0 40 36 31 29 37 31 33 40 32 27 23 32 28 28 25
AU-Yao	40 0 38 45 30 38 36 42 49 44 31 32 40 26 36 29
AU-Kharia	36 38 0 47 28 36 45 41 26 28 34 41 47 40 45 38
AU-Khasi	31 45 47 0 40 51 46 52 38 49 38 45 50 31 39 33
AU-Cambodian	29 30 28 40 0 65 55 51 46 56 54 44 57 49 52 50
AU-Nancowry	37 38 36 51 65 0 67 55 46 57 53 49 58 50 57 49
AU-Temiar	31 36 45 46 55 67 0 60 41 49 58 56 66 58 78 59
AU-Bru	33 42 41 52 51 55 60 0 43 50 54 50 65 46 58 50
AU-Lakkia	40 49 26 38 46 46 41 43 0 48 42 30 38 38 41 38
AU-Thai	32 44 28 49 56 57 49 50 48 0 43 39 48 40 46 38

AU-Atayal	27	31	34	38	54	53	58	54	42	43	0	66	60	58	65	75
AU-WBukidnon	23	32	41	45	44	49	56	50	30	39	66	0	62	47	60	57
AU-Mamay	32	40	47	50	57	58	66	65	38	48	60	62	0	57	68	56
AU-Mor2	28	26	40	31	49	50	58	46	38	40	58	47	57	0	69	63
AU-Arosi	28	36	45	39	52	57	78	58	41	46	65	60	68	69	0	72
AU-Easter	25	29	38	33	50	49	59	50	38	38	75	57	56	63	72	0

Langues plus proches que le plus proche voisin

AU-Miao	ST-Tibetan (45)	ST-Lahu (46)
AU-Yao		
AU-Kharia	AL-Khalkha (48)	AL-Evenki (53) AM-Carib (54) DR-Malto (49)
	IE-Hindi (63)	IP-Andamanese (55) IP-Kunimaipa (50) IP-Woisika (53) IP-Savosavo (55) ST-Lahu (49) UR-Yukaghir (48)
AU-Khasi	IP-Andamanese (54)	IP-Kwoma (55) NS-Kanuri (53) ST-Khami (53) ST-Lahu (54)
AU-Cambodian		
AU-Nancowry		
AU-Temiar		
AU-Bru		
AU-Lakkia		
AU-Thai		
AU-Atayal		
AU-WBukidnon		
AU-Mamay	NK-Fula (70)	NK-Banda (69) NS-Kanuri (72)
AU-Mor2		
AU-Arosi		
AU-Easter		

Famille 6, nb. d'elements 3

CA-Georgian	0	66	56
CA-Kabardian	66	0	57
CA-Xnalug	56	57	0

Langues plus proches que le plus proche voisin

CA-Georgian	IE-Armenian (78)
CA-Kabardian	
CA-Xnalug	

Famille 7, nb. d'elements 4

DR-Brahui	0	53	59	40
DR-Malto	53	0	61	33
DR-Tamil	59	61	0	32
DR-Elamite	40	33	32	0

Langues plus proches que le plus proche voisin

DR-Brahui	
DR-Malto	
DR-Tamil	
DR-Elamite	AA-Walamo (41) AA-Afar (41) AL-Bashkir (46) AL-Evenki (44) AM-Mojave (46) AS-Maranungu (46) AS-WikNunkan (45) AS-Pitjantjatj (45) AU-Cambodian (47) AU-Nancowry (41) AU-Atayal (44) AU-Mamay (51) AU-Mor2 (47) EA-Aleut (46) EA-Yupik (43) IE-Gaelic (42) IE-Rumanian (43) IP-Andamanese (43) IP-Kate (47) IP-Kunimaipa (45) IP-Waskia (58) IP-Woisika (46) NK-Moro (45) NK-Loma (46) NK-Fula (44) NK-More (47) NK-Banda (43) NK-Shona (43) NS-Songhai (52) NS-Fur (52)

NS-Kanuri (61) NS-Berta (51) NS-Anuak (43) NS-Bagirma (45)
 PS-Chukchi (47) PS-Kamchadal (43) ST-Khami (43) ST-Bodo (51)
 UR-Yukaghir (43) UR-Yenets (47)

Famille 8, nb. d'elements 2

EA-Aleut 0 77
 EA-Yupik 77 0

Langues plus proches que le plus proche voisin
 EA-Aleut
 EA-Yupik

Famille 9, nb. d'elements 7

IE-Albanian 0 43 63 57 57 79 36
 IE-Armenian 43 0 47 43 44 42 44
 IE-Russian 63 47 0 67 61 67 42
 IE-Gaelic 57 43 67 0 52 66 40
 IE-English 57 44 61 52 0 58 41
 IE-Rumanian 79 42 67 66 58 0 37
 IE-Hindi 36 44 42 40 41 37 0

Langues plus proches que le plus proche voisin

IE-Albanian
 IE-Armenian AA-Tera (50) AA-Walamo (51) AL-Khalkha (52) CA-Georgian (78)
 CA-Kabardian (61) CA-Xnalug (53) DR-Brahui (48) ST-Bodo (48)
 ST-Lahu (50)

IE-Russian
 IE-Gaelic
 IE-English
 IE-Rumanian
 IE-Hindi AA-Walamo (45) AA-Afar (47) AL-Khalkha (45) AL-Evenki (47)
 AU-Kharia (63) CA-Georgian (47) DR-Brahui (57) DR-Malto (46)
 DR-Tamil (53) IP-Andamanese (45) ST-Bodo (46) ST-Lahu (46)
 UR-Yukaghir (45)

Famille 10, nb. d'elements 8

IP-Andamanese 0 52 49 65 50 54 53 35
 IP-Kate 52 0 62 63 63 46 46 47
 IP-Kunimaipa 49 62 0 63 64 38 58 51
 IP-Waskia 65 63 63 0 72 44 46 48
 IP-Woisika 50 63 64 72 0 44 53 60
 IP-Kwoma 54 46 38 44 44 0 50 32
 IP-Savosavo 53 46 58 46 53 50 0 46
 IP-Reefs 35 47 51 48 60 32 46 0

Langues plus proches que le plus proche voisin

IP-Andamanese AL-Evenki (68) NS-Fur (67) PS-Chukchi (67) ST-Bodo (66)

IP-Kate
 IP-Kunimaipa
 IP-Waskia
 IP-Woisika
 IP-Kwoma AA-Walamo (58) AL-Japanese (55) AU-Khasi (55)
 IP-Savosavo
 IP-Reefs AU-Arosi (69)

Famille 11, nb. d'elements 3

KH-Kung	0	36	37
KH-Korana	36	0	23
KH-Hua	37	23	0

Langues plus proches que le plus proche voisin

KH-Kung	
KH-Korana	AM-Itonama (37) AS-Djingili (43) ST-Bodo (37)
KH-Hua	AA-Tera (46) AA-Walamo (38) AA-Afar (38) AL-Bashkir (40) AL-Khalkha (40) AL-Evenki (40) AM-Mojave (50) AM-Mixtec (38) AM-Carib (41) AM-Tucuna (39) AS-WikNunkan (41) AU-Yao (38) AU-Cambodian (45) AU-Nancowry (42) AU-Temiar (49) AU-Bru (39) AU-Lakkia (40) AU-Thai (43) AU-Atayal (45) AU-WBukidnon (38) AU-Mamay (48) AU-Mor2 (45) AU-Arosi (46) AU-Easter (43) CA-Georgian (38) IE-Armenian (41) IE-Russian (38) IE-Gaelic (43) IE-English (42) IE-Rumanian (38) IP-Andamanese (40) IP-Kate (47) IP-Kunimaipa (46) IP-Waskia (54) IP-Woisika (47) IP-Reefs (43) ND-Carrier (40) ND-Navajo (39) NK-Moro (38) NK-Loma (46) NK-Fula (47) NK-More (57) NK-Banda (42) NK-Awing (43) NK-Shona (46) NS-Songhai (56) NS-Fur (49) NS-Kanuri (49) NS-Berta (56) NS-Anuak (44) NS-Bagirma (46) ST-Sgaw (45) ST-Tibetan (38) ST-Khami (46) ST-Bodo (45) ST-Lahu (40) UR-Yukaghir (42) UR-Yenets (38) UR-Finnish (39)

Famille 12, nb. d'elements 3

ND-Tlingit	0	61	44
ND-Carrier	61	0	62
ND-Navajo	44	62	0

Langues plus proches que le plus proche voisin

ND-Tlingit	
ND-Carrier	
ND-Navajo	

Famille 13, nb. d'elements 7

NK-Moro	0	47	68	50	59	61	61
NK-Loma	47	0	51	64	58	52	48
NK-Fula	68	51	0	57	67	58	72
NK-More	50	64	57	0	53	53	58
NK-Banda	59	58	67	53	0	60	60
NK-Awing	61	52	58	53	60	0	59
NK-Shona	61	48	72	58	60	59	0

Langues plus proches que le plus proche voisin

NK-Moro	
NK-Loma	
NK-Fula	
NK-More	NS-Songhai (77) NS-Fur (65)
NK-Banda	AU-Mamay (69)
NK-Awing	
NK-Shona	

Famille 14, nb. d'elements 6

NS-Songhai	0	71	63	68	55	62
NS-Fur	71	0	76	60	53	60
NS-Kanuri	63	76	0	67	58	62
NS-Berta	68	60	67	0	55	59
NS-Anuak	55	53	58	55	0	60
NS-Bagirma	62	60	62	59	60	0

Langues plus proches que le plus proche voisin

NS-Songhai	NK-More (77)
NS-Fur	ST-Bodo (78)
NS-Kanuri	
NS-Berta	
NS-Anuak	AU-Mamay (66) NK-Moro (62) NK-Fula (64) NK-Banda (61)
NS-Bagirma	

Famille 15, nb. d'elements 2

PS-Chukchi	0	62
PS-Kamchadal	62	0

Langues plus proches que le plus proche voisin

PS-Chukchi	IP-Andamanese (67) UR-Yukaghir (63)
PS-Kamchadal	

Famille 16, nb. d'elements 6

ST-Mandarin	0	50	37	43	50	45
ST-Sgaw	50	0	42	52	49	49
ST-Tibetan	37	42	0	53	36	52
ST-Khami	43	52	53	0	58	65
ST-Bodo	50	49	36	58	0	51
ST-Lahu	45	49	52	65	51	0

Langues plus proches que le plus proche voisin

ST-Mandarin	
ST-Sgaw	AM-Mojave (53) AU-Cambodian (54) AU-Nancowry (53)
ST-Tibetan	
ST-Khami	
ST-Bodo	AA-Afar (62) AL-Bashkir (64) AL-Khalkha (59) AL-Evenki (59) AM-Carib (65) AS-Banjalong (59) IE-Gaelic (60) IP-Andamanese (66) IP-Waskia (69) NK-Loma (61) NK-More (61) NS-Songhai (64) NS-Fur (78) NS-Kanuri (60) PS-Chukchi (61) UR-Yukaghir (66)
ST-Lahu	

Famille 17, nb. d'elements 3

UR-Yukaghir	0	61	57
UR-Yenets	61	0	51
UR-Finnish	57	51	0

Langues plus proches que le plus proche voisin

UR-Yukaghir	AA-Afar (62) AL-Bashkir (66) AM-Carib (71) IP-Andamanese (65) IP-Waskia (65) NS-Fur (64) PS-Chukchi (63) ST-Bodo (66)
UR-Yenets	
UR-Finnish	

Annexe III : Similitude moyenne aux classes

			AA	AL	AM	AS	AU	CA	DR	EA	IE	IP	KH	ND
NK	NS	PS	ST	UR	IS									
AA-Tuareg			40.3	29.8	33.3	28.3	33.3	26.3	34.3	28.0	40.6	32.1	29.3	23.0
39.7	39.7	29.5	28.5	31.0	31.8	*								
AA-Tera			43.3	47.0	37.9	30.5	39.2	40.0	40.0	47.0	49.4	44.4	34.0	38.0
49.7	54.0	50.0	42.8	47.7	46.3	*								
AA-Walamo			50.8	52.6	41.9	31.8	35.8	47.3	47.5	39.5	50.4	45.9	32.7	46.3
41.7	45.7	51.0	42.2	49.7	51.8	*								
AA-Afar			51.5	51.6	40.9	39.7	34.9	41.0	52.0	44.0	49.7	49.5	30.3	34.7
42.6	45.8	48.5	41.5	55.0	51.7	*								
AA-Iraqi			47.8	35.6	35.8	29.5	35.2	32.0	39.3	42.5	47.7	33.1	28.0	32.0
38.9	41.2	39.0	32.7	36.7	35.7									
AL-Bashkir			47.0	52.5	36.5	34.0	37.4	41.3	45.8	53.5	44.0	48.0	30.7	41.7
41.9	47.8	52.5	46.5	58.3	53.0	*								
AL-Khalkha			43.8	53.0	38.4	35.5	38.8	42.3	47.0	45.5	46.6	48.3	31.7	39.7
43.9	46.2	48.5	42.2	53.7	53.5	*								
AL-Evenki			44.4	54.8	38.5	39.5	40.1	38.0	50.8	47.5	44.4	50.9	30.7	34.3
46.3	50.2	53.0	42.2	57.7	54.7	*								
AL-Korean			34.4	44.0	36.6	31.2	38.1	42.0	34.5	40.0	34.4	40.3	29.0	46.0
31.7	34.8	52.5	46.3	48.7	47.3	*								
AL-Japanese			47.0	47.3	37.6	30.5	33.4	36.7	47.5	34.0	44.0	42.5	29.7	36.7
34.9	42.3	42.0	39.7	49.3	47.7	*								
AM-Halkomelen			34.2	28.4	29.0	20.8	26.5	35.3	27.5	28.5	34.0	21.1	24.3	32.3
25.1	26.5	31.5	26.0	28.0	26.3	*								
AM-Tuscarora			31.4	29.4	34.9	30.5	29.9	25.7	31.0	34.5	29.1	30.8	27.7	32.7
28.7	29.0	31.0	27.2	31.0	30.7									
AM-Mojave			38.0	43.6	35.6	36.0	42.1	30.7	38.8	42.5	36.7	45.9	32.0	42.3
45.4	50.5	49.5	48.8	45.7	45.3	*								
AM-Jacaletc			33.4	32.8	34.4	27.0	39.1	42.7	29.3	32.0	34.3	32.5	27.7	38.7
37.6	36.3	38.0	35.3	30.3	36.0	*								
AM-Mixtec			38.8	33.6	38.7	32.0	41.6	25.0	35.0	24.5	39.7	39.0	30.3	27.0
46.0	42.2	30.0	32.8	31.7	34.8	*								
AM-Itonama			41.8	32.8	38.6	31.5	36.5	29.7	35.0	29.5	41.9	34.8	30.3	29.7
38.4	35.3	34.5	30.3	34.7	35.8	*								
AM-Carib			45.8	58.6	40.0	41.2	45.3	38.0	49.8	48.0	42.3	55.3	33.3	39.0
46.3	49.2	52.0	44.8	61.3	56.5	*								
AM-Tucuna			40.0	41.0	38.6	34.5	40.2	28.7	34.8	33.5	37.6	44.8	31.3	33.0
43.6	46.0	39.5	40.0	37.3	41.8	*								
AS-Tiwi			35.8	35.6	33.5	37.4	33.8	24.0	38.8	44.0	34.1	37.3	23.7	24.0
35.6	32.7	43.5	33.2	38.0	36.0	*								
AS-Djingili			24.2	19.8	22.0	30.0	19.6	14.0	27.3	24.5	22.9	22.1	27.0	15.3
22.0	25.0	22.5	19.3	21.7	20.3									
AS-Maranungu			30.2	31.8	32.1	46.4	36.4	19.7	35.8	35.5	27.7	43.1	21.7	22.0
37.0	37.7	33.5	32.5	31.7	30.8									
AS-WikNunkan			29.0	37.0	35.6	48.6	44.8	22.7	36.3	40.5	25.9	49.8	28.0	29.7
41.9	46.2	40.5	38.8	40.0	36.8	*								
AS-Banjalang			46.6	46.0	36.4	36.2	30.4	35.3	43.5	41.5	42.1	45.1	27.0	28.7
36.7	42.0	44.0	37.2	48.7	44.2	*								
AS-Pitjantjatj			26.0	34.6	30.5	48.2	37.2	21.0	37.8	43.0	24.7	45.4	23.3	27.0
35.6	39.3	35.0	36.3	36.3	34.3									

			AA	AL	AM	AS	AU	CA	DR	EA	IE	IP	KH	ND
NK	NS	PS	ST	UR	IS									
AU-Miao			26.8	28.0	25.6	21.3	31.5	27.0	25.5	31.0	32.1	30.5	24.7	31.7
31.3	31.2	33.5	38.5	29.3	34.0	*								
AU-Yao			34.4	35.4	33.3	22.7	37.1	32.7	26.8	26.5	36.1	33.8	32.0	33.3
35.6	36.7	33.5	39.7	36.0	36.8	*								
AU-Kharria			36.8	46.6	36.9	37.8	38.0	38.0	43.0	41.5	39.7	47.6	31.7	35.3
39.0	40.8	42.5	42.3	43.7	54.8	*								
AU-Khasi			42.4	45.2	38.5	31.8	42.3	34.7	34.8	34.0	41.3	40.8	32.3	35.
39.0	43.0	44.0	47.8	42.7	46.5	*								
AU-Cambodian			34.2	33.0	34.5	33.8	47.1	22.7	31.5	38.0	36.4	38.4	28.3	29.3
46.7	52.5	39.5	41.0	40.0	33.8	*								
AU-Nancowry			37.0	36.6	37.6	35.3	51.2	25.0	31.0	48.0	39.3	44.0	28.3	34.7
47.6	49.2	44.5	46.2	42.3	38.8									
AU-Temiar			36.2	37.8	39.6	39.5	53.7	25.0	34.0	37.0	37.0	47.5	32.7	30.0
51.0	50.5	36.0	40.8	41.0	37.7									
AU-Bru			33.6	38.2	39.3	31.0	50.0	26.0	31.3	32.0	36.0	39.6	30.0	30.0
46.3	44.7	38.0	42.2	38.3	39.2									
AU-Lakkia			28.2	26.0	31.3	25.2	40.3	23.0	22.8	27.5	30.0	30.6	28.0	31.7
39.0	39.5	32.5	38.8	30.7	28.2									
AU-Thai			34.0	31.6	35.3	29.2	44.5	26.0	27.5	29.5	38.4	36.6	31.7	27.3
43.9	47.8	32.5	43.3	35.0	32.5	*								
AU-Atayal			36.8	39.8	41.5	37.3	50.5	26.7	37.5	43.0	38.9	42.5	31.0	33.7
46.6	44.8	44.0	39.7	44.0	39.0									
AU-WBukidnon			39.0	44.6	39.1	35.7	46.7	30.0	40.0	43.0	40.9	43.1	27.7	28.3
45.3	41.8	43.5	39.0	45.7	41.7									
AU-Mamay			47.2	48.0	45.6	36.8	53.6	34.0	44.0	40.0	50.9	49.0	35.3	36.3
60.7	61.0	46.5	44.2	45.3	47.8	*								
AU-Mor2			34.0	34.2	40.0	40.3	46.7	25.0	35.0	43.5	34.1	45.8	31.7	28.7
51.0	46.3	39.0	37.3	39.0	35.3	*								
AU-Arosi			37.4	39.2	42.8	40.5	54.3	25.7	35.5	40.5	37.1	51.5	31.7	31.3
50.3	49.7	37.0	38.7	44.0	38.7									
AU-Easter			32.8	37.0	41.5	40.7	48.8	23.0	34.5	36.0	34.3	43.9	29.3	30.0
42.7	41.7	40.0	37.5	40.3	37.7									
CA-Georgian			40.0	47.8	33.5	25.0	30.8	61.0	39.0	38.5	45.6	38.3	31.7	47.0
33.9	37.2	42.5	39.7	42.3	47.3									
CA-Kabardian			35.0	35.6	29.1	21.5	26.4	61.5	30.8	36.0	36.3	30.4	28.7	46.7
31.6	36.2	36.0	35.3	29.3	35.8									
CA-Xnalug			37.0	36.8	33.3	21.8	26.2	56.5	34.5	28.5	37.6	28.8	27.7	39.3
29.6	29.2	35.0	31.7	32.7	37.2									
DR-Brahui			46.6	47.8	33.5	33.3	32.1	41.7	50.7	40.5	47.6	41.9	29.7	35.3
41.7	40.8	45.5	37.2	44.3	48.3									
DR-Malto			44.6	46.6	39.4	39.3	33.5	33.7	49.0	37.5	42.1	43.1	28.3	31.0
39.1	40.3	41.5	35.7	48.0	47.2									
DR-Tamil			42.0	46.4	36.0	35.8	31.2	34.7	50.7	33.5	45.1	39.6	27.0	30.7
36.7	35.7	37.5	33.8	44.0	45.7									
DR-Elamite			37.2	39.6	31.6	37.7	36.8	29.0	35.0	44.5	37.1	42.3	24.0	31.7
44.0	50.7	45.0	38.3	42.3	39.3	*								
EA-Aleut			40.0	43.2	34.3	38.0	36.3	35.7	38.3	77.0	34.6	44.3	25.3	43.3
35.6	40.3	55.5	42.3	48.3	45.5									
EA-Yupik			40.4	45.0	34.0	38.3	37.6	33.0	39.8	77.0	38.4	44.9	26.7	39.3
36.7	42.0	58.5	44.0	55.7	46.2									

			AA	AL	AM	AS	AU	CA	DR	EA	IE	IP	KH	ND
NK	NS	PS	ST	UR	IS									
IE-Albanian			48.6	39.0	37.6	28.7	38.2	34.3	40.8	32.0	55.8	36.8	31.3	32.7
46.9	45.2	36.5	36.3	39.3	38.8									
IE-Armenian			41.8	45.6	31.9	23.5	32.7	64.0	37.3	36.0	43.8	36.8	32.7	41.3
36.3	38.5	42.5	42.0	40.3	46.2	*								
IE-Russian			52.0	44.8	39.8	31.2	37.5	36.3	46.0	38.0	57.8	41.8	32.3	33.3
46.7	44.7	44.5	41.2	46.7	45.3									
IE-Gaelic			51.4	40.8	42.0	33.8	42.8	33.3	43.8	40.0	54.2	46.5	33.7	31.0
48.9	48.7	37.0	42.7	48.3	44.5									
IE-English			48.4	45.8	35.4	29.8	37.3	36.7	42.5	38.0	52.2	39.4	32.0	34.3
40.7	43.8	41.5	38.7	44.7	41.8									
IE-Rumanian			52.4	39.8	41.4	30.3	43.1	34.7	42.8	35.0	58.2	41.1	33.7	32.3
52.6	49.8	39.0	40.3	40.0	42.0									
IE-Hindi			38.4	43.0	30.6	29.7	32.1	39.3	48.0	36.5	40.0	36.5	29.7	35.7
34.3	36.3	39.0	40.0	42.3	49.5	*								
IP-Andamanese			46.8	59.0	39.0	41.0	41.3	37.3	45.8	48.0	44.3	51.1	30.7	38.3
47.3	53.3	58.5	48.8	57.0	57.7	*								
IP-Kate			40.4	43.0	37.1	37.0	43.0	33.7	38.8	50.5	40.0	54.1	34.0	40.7
52.9	53.2	45.0	48.0	49.7	46.5									
IP-Kunimaipa			33.6	43.4	34.6	39.7	38.3	30.3	42.8	48.5	36.0	55.0	32.0	32.7
47.6	46.5	40.0	43.3	46.0	45.5									
IP-Waskia			41.0	50.4	39.5	43.7	43.8	33.3	43.5	46.0	39.4	57.3	34.3	36.0
50.1	57.7	48.0	52.8	53.0	51.2	*								
IP-Woisika			40.8	46.6	41.1	47.7	47.4	32.3	44.8	44.0	37.7	58.0	32.7	34.7
49.9	54.7	41.5	47.3	50.7	48.5									
IP-Kwoma			48.0	45.8	38.4	31.3	36.3	35.3	38.0	40.0	41.0	44.0	29.7	37.7
38.4	39.5	41.5	40.3	42.0	46.7	*								
IP-Savosavo			44.0	46.4	37.8	44.0	36.4	34.0	47.3	45.0	43.1	50.3	28.3	29.3
43.6	41.2	44.5	39.0	48.3	47.0									
IP-Reefs			33.4	33.2	36.5	39.3	46.2	23.3	33.0	34.5	37.0	45.6	30.0	26.0
51.3	46.8	30.0	35.8	39.0	32.5	*								
KH-Kung			25.8	25.2	25.9	20.2	28.4	30.3	22.8	19.0	29.9	24.9	36.5	29.3
30.0	31.8	23.0	27.2	22.7	25.8									
KH-Korana			30.2	27.4	27.0	25.0	21.4	23.7	27.0	24.0	28.0	26.4	29.5	22.3
23.1	23.8	27.0	26.2	27.7	29.3	*								
KH-Hua			36.6	38.4	36.0	30.2	41.4	34.0	32.0	35.0	38.7	43.1	30.0	38.3
45.6	50.0	36.0	41.8	39.7	36.8	*								
NK-Moro			41.8	40.0	39.0	37.7	44.7	29.3	44.0	37.0	43.7	47.6	29.7	30.0
57.7	59.2	41.5	39.2	39.3	41.7	*								
NK-Loma			40.8	45.6	38.6	37.5	40.9	35.7	43.0	41.0	38.7	54.8	33.7	35.3
53.3	54.7	45.0	47.7	44.0	46.7	*								
NK-Fula			44.4	39.6	42.0	38.7	49.9	29.3	42.0	36.5	45.0	47.8	33.7	29.0
62.2	57.7	41.0	38.3	40.0	40.3									
NK-More			46.2	42.4	39.5	38.0	43.1	35.7	45.0	41.0	47.1	51.0	38.0	35.0
55.8	61.7	42.5	48.0	46.3	45.7	*								
NK-Banda			42.8	41.2	39.3	30.7	47.9	32.7	39.8	34.0	45.4	45.4	31.3	28.3
59.5	56.5	41.5	40.7	39.3	41.5									
NK-Awing			37.2	32.2	36.9	30.3	43.3	28.3	33.0	30.0	39.4	42.8	30.0	28.7
57.2	52.3	34.5	37.7	33.0	34.0									
NK-Shona			44.4	37.0	37.0	30.7	43.4	30.7	36.0	33.5	46.9	44.1	34.0	31.7
59.7	57.3	39.0	39.3	37.3	38.2									

			AA	AL	AM	AS	AU	CA	DR	EA	IE	IP	KH	ND
NK	NS	PS	ST	UR	IS									
ND-Tlingit			34.8	37.8	32.1	22.2	27.0	42.3	31.5	38.5	31.7	32.9	27.3	52.5
30.3	33.0	40.5	32.8	33.0	37.3									
ND-Carrier			37.4	43.0	37.8	26.8	35.1	48.0	34.0	45.0	37.1	36.1	31.3	61.5
34.1	39.3	48.0	42.8	39.3	43.2									
ND-Navajo			32.2	38.2	33.1	24.3	33.0	42.7	31.0	40.5	34.3	34.3	31.3	53.0
29.0	33.7	41.0	42.5	35.7	39.5									
NS-Songhai			46.8	45.2	40.0	38.8	42.3	35.0	43.5	44.5	42.4	52.8	37.7	36.7
58.1	63.8	48.5	47.0	49.3	46.2									
NS-Fur			49.4	53.4	41.6	40.8	43.3	41.0	46.5	48.0	48.0	56.5	36.0	40.7
55.7	64.0	51.0	53.8	54.3	53.7									
NS-Kanuri			49.6	48.6	41.5	36.7	48.1	38.0	46.8	40.5	48.1	50.5	36.3	39.3
58.7	65.2	48.5	48.2	47.7	48.5									
NS-Berta			48.0	44.2	39.1	33.8	42.9	40.7	42.8	43.0	45.9	45.6	38.3	41.0
55.4	61.8	46.5	42.3	44.0	41.8									
NS-Anuak			37.4	38.2	38.6	38.2	46.9	25.3	35.5	33.5	37.6	45.0	31.0	27.3
57.6	56.2	38.0	37.2	37.3	37.5	*								
NS-Bagirma			40.4	36.0	35.4	34.5	46.9	25.0	36.3	37.5	41.1	44.3	32.0	27.0
56.7	60.6	37.0	43.2	41.3	37.8									
PS-Chukchi			44.6	52.8	39.9	40.0	40.1	33.7	44.8	57.0	40.3	46.8	27.7	42.3
40.9	46.2	62.0	49.5	57.3	54.5									
PS-Kamchadal			42.6	46.6	36.6	33.0	38.2	42.0	40.0	57.0	39.7	40.5	29.7	44.0
40.6	43.7	62.0	43.7	47.0	48.8									
ST-Mandarin			33.0	39.8	31.1	30.3	36.6	31.0	31.5	44.5	37.1	34.9	28.3	35.0
32.6	36.5	46.5	45.0	41.7	40.7	*								
ST-Sgaw			35.6	36.4	36.5	30.7	44.3	30.3	31.0	41.0	35.6	39.0	31.3	35.0
43.1	45.5	45.5	48.4	36.3	39.2									
ST-Tibetan			30.2	35.0	31.9	31.0	36.9	30.3	30.0	35.0	32.3	37.0	29.3	39.0
34.0	37.5	37.0	44.0	34.0	40.5									
ST-Khami			37.6	46.4	37.3	33.5	45.0	40.0	38.5	41.0	41.4	52.0	34.3	41.3
46.6	49.3	47.5	54.2	46.3	51.0									
ST-Bodo			50.2	55.8	42.9	41.7	41.9	40.7	49.5	52.5	51.1	55.5	36.0	42.7
51.3	58.3	55.0	48.8	57.3	56.0	*								
ST-Lahu			38.6	46.8	34.4	30.2	41.8	41.0	37.0	45.0	43.4	48.3	31.0	43.3
41.7	44.5	48.0	52.4	44.3	52.7	*								
UR-Yukaghir			47.2	57.0	41.3	39.7	42.4	37.0	47.0	54.5	43.9	52.8	31.0	41.0
41.1	49.5	55.5	49.2	59.0	57.0									
UR-Yenets			43.2	53.4	36.8	37.0	39.9	34.0	44.5	56.0	41.3	50.1	30.3	37.0
40.3	44.8	54.5	42.0	56.0	50.2									
UR-Finnish			41.6	50.2	34.5	31.5	37.2	33.3	42.5	45.5	44.1	41.8	28.7	30.0
38.3	42.7	46.5	38.8	54.0	44.2									