



République du Sénégal
Ministère de l'Enseignement Supérieur, de la Recherche et de L'innovation
UNIVERSITÉ CHEIKH ANTA DIOP de DAKAR
Centre de Linguistique Appliquée de Dakar

FICHE TECHNIQUE

Construction et structuration d'une base de données génétiques sur les langues

Language	Dialect	Classification	Country	Area	Continent	Alternate name	Population	GLD number	Renseignement	Consonnes																											
										β	ɱ	ɸ	θ	ð	tʰ	pʰ	h	ʒ	ʁ	l	ʎ	ʝ	ʂ	ʃ	ʣ	ʤ	ʥ	ʦ	ʧ	ʨ	ʩ	ʪ	ʫ	ʬ			
1																																					
2	Proto-Kartvelian	Kartvelian	Georgia		Europe			1595																													
3	Georgian		Georgia		Europe			1596																													
4	Svan		Georgia	North-West	Europe		35000	1597																													
5	Mingrelian		Georgia	West	Europe	Zan	300000	1598																													
6	Laz		Georgia	North-West	Europe		50000	1599	n.r.																												
7	Basque		Spain	North-West	Europe		1 million	1600																													
8	Ossetian		SW Russia	Abxar ASSR	Europe		85000	1601	Extinct (1992)																												
9	Abkhaz		SW Russia	Abxar ASSR	Europe	Abkhaz	55000	1602																													
10	Abaza		SW Russia	Abxar ASSR	Europe		30000	1603																													
11	Abaza	Tapanta	SW Russia	Abxar ASSR	Europe		85000	1604																													
12	Adygh		SW Russia	Adygh AO	Europe	Adygh	80000	1605																													
13	Kabardian		SW Russia	Abxar ASSR	Europe		270000	1606																													
14	Bats		Georgia		Europe	Tsova-Tush	3000	1607																													
15	Chechen		SW Russia	Igush ASSR	Europe		420000	1608																													
16	Igush		SW Russia	Igush ASSR	Europe		180000	1609																													
17	Avar		SW Russia	estan ASSR	Europe		270000	1610																													
18	Andi		SW Russia	estan ASSR	Europe		9000	1611																													
19	Botl		SW Russia	estan ASSR	Europe		3000	1612																													
20	Godoberi		SW Russia	estan ASSR	Europe		2500	1613																													
21	Chamalal		SW Russia	estan ASSR	Europe		4000	1614																													
22	Bagulal		SW Russia	estan ASSR	Europe		4000	1615																													
23	Tind		SW Russia	estan ASSR	Europe		5000	1616																													
24	Karata		SW Russia	estan ASSR	Europe		5000	1617																													
25	Acvax		SW Russia	estan ASSR	Europe		5000	1618																													
26	Xvarshi	Inxokari	SW Russia	estan ASSR	Europe		1800	1619																													
27	Dido		SW Russia	estan ASSR	Europe		7000	1620																													
28	Hinux		SW Russia	estan ASSR	Europe	Tsez	200	1621																													
29	Bezhta		SW Russia	estan ASSR	Europe		2500	1622																													
30	Bezhta		SW Russia	estan ASSR	Europe		2500	1623																													
31	Hunzib	Tjadal	SW Russia	estan ASSR	Europe		600	1624																													
32	Lak		SW Russia	estan ASSR	Europe		64000	1625																													
33	Dargwa		SW Russia	estan ASSR	Europe		160000	1626																													
34	Dargwa	Kubachi	SW Russia	estan ASSR	Europe		160000	1627																													
35	Archi		SW Russia	estan ASSR	Europe		1000	1628																													
36	Xinalu		SW Russia	estan ASSR	Europe		1000	1629																													

Conçue par :

Dr Augustin NDIONE

Chargé de recherche en linguistique descriptive et didactique des langues

Dr Noël Bernard BIAGUI

Chargé de recherche en linguistique descriptive et didactique des langues

JUIN 2022

Résumé : A travers le globe, il existe un très grand nombre de langues (cf. ethnologue.com 7117 langues, <https://www.ethnologue.com>). Elles sont différentes les unes des autres, cependant les linguistes pensent qu'au-delà de ces différences les langues ont nécessairement quelque chose en commun. Dans le cadre de la recherche et les travaux s'inscrivant en linguistique générale et plus précisément dans la comparaison des langues. Nous avons pu nous intéresser aux caractéristiques des langues et créer des bases de comparaison tant génétique que typologique de langues parlées dans le monde. En effet, comparer les langues permet de mettre en lumière ce que les langues ont en commun et permet également d'inférer quelque chose sur la faculté de langage.

Plan de la fiche technique

<i>Plan de la fiche technique</i>	2
<i>Introduction</i>	2
1. Matériel et Méthodes	2
2. Productions et résultats	5
3. Exploitation possible à partir de la base de données aréales	7
<i>Conclusion</i>	9
<i>Références bibliographiques</i>	9

Introduction

Les langues peuvent avoir une origine commune (voir A. Meillet), elles peuvent aussi avoir une forme commune (voir J. Greenberg), et des principes inhérents, comme la créativité, permettent au sujet de produire un nombre illimité de langues à partir d'une forme linguistique en nombre limité (voir N. Chomsky), ou encore des opérations visant à gérer l'activité linguistique (voir A. Culioli). En fait, comparer les langues est un excellent moyen d'essayer de comprendre en quoi les langues ont quelque chose en commun et en quoi elles diffèrent. Ces propos de F. de Saussure (CLG, p.20) abondent dans le même sens : « la tâche de la linguistique consiste entre autres à faire la description et l'histoire de toutes les langues qu'elle pourra atteindre, ce qui revient à faire l'histoire des familles de langues et à reconstituer dans la mesure du possible les langues mères de chaque famille... ». Notre objectif dans cette opération de recherche a été de constituer une large base de données sur les langues afin de pouvoir identifier des spécificités géographiques en termes de caractéristiques génétiques ou en termes de caractéristiques typologiques.

1. Matériel et Méthodes

Les langues ont tendance à évoluer et à varier, ce faisant, certaines langues changent et d'autres se transforment. Ce genre de recueil permet de mettre en lumière les caractéristiques des langues et cela permet aussi de mettre en avant les points de variations d'une langue à l'autre compte tenu de leur origine commune.

Comment les langues ont-elles été choisies ?

A priori, toutes les langues du monde étaient éligibles, toutefois, le seul critère retenu par l'équipe l'accessibilité d'informations sur la langue en question. Le critère géographique a prévalu lors du choix des langues, ainsi, les langues ont été regroupées par continent au départ.

Il s'est agi dans ce travail de collecter le plus d'informations possible sur les langues à travers le globe. Ainsi il était question de mettre en place une équipe chargée de recueillir les informations pertinentes. Pour ce faire une grille a été élaborée.

Il faut noter que deux tableurs ont été réalisés, une grille pour les consonnes et une grille pour les voyelles. Ce faisant, un travail exhaustif a été proposé tant pour les voyelles que pour les consonnes.

a) Grille de collectes de données (Grille vierge)

IDENTIFICATION DE LA LANGUE		
Nom :		
Lieu :		
Population :		
Classification :		
Consonnes		
Voyelles		
Accent		
Ton		

b) Exemples de grilles complétées

Les grilles ainsi complétées ont ensuite été intégrées à la base de données créées sous Excel.

IDENTIFICATION DE LA LANGUE		
Nom :	Français	Dialecte : Parisien
Lieu :	France	
Population :	66 millions	Compte non tenu des francophones
Classification :	Indo-Hittite: Indo-European: Italic: Romance: Western: Gallo-Romance	
Consonnes	Occlusives: p b t d k g Fricatives: f v s z š ž Nasales: m n ñ Latérales: l Vibrantes: r Glides: j ɥ w	
Voyelles	Antérieures: i y e ø ε œ Centrales: ə a Postérieures: u o ɔ (ɑ) Nasales: (œ) œ ò ã	
Accent	Non pertinent	
Ton	Non pertinent	

IDENTIFICATION DE LA LANGUE		
Nom :	Dyirbal	
Lieu :	NE Australia (NE Queensland)	
Population :	40	
Classification :	Australian: Pama-Nyungan: Dyirbalic	
Consonnes	Occlusives: b d ɟ g Nasales: m n ñ ŋ Latérales: l Vibrantes: r ɹ Glides: j w	
Voyelles	Antérieures: i Centrales: a Postérieures: u	
Accent	initiale	
Ton	Non pertinent	

IDENTIFICATION DE LA LANGUE		
Nom :	Nama	Autre nom: Khoekhoe
Lieu :	Afrique du Sud	
Population :	234.000	
Classification :	Khoisan: Central: Nama	
Consonnes	Occlusives: p tʰ (c) k ʔ Affriquées: tʃʰ kʰ Fricatives: (f) (v) s x h Nasales: m n	

	Latérales: (l) Vibrantes: r Clicks: /h ʔ /x n/ ɛ ɛ ^h ɛ ^ʔ ɛ ^x nɛ ! ʔ ^h ! ^ʔ ! ^x n! ^h ^ʔ ^x n	
Voyelles	Antérieures: i e Centrales: (ə) a Postérieures: u o Nasales: ã ü Diphthongues: ai ae ao au oe oa ui ā āū ō ā ū	
Accent	Non pertinent	
Ton	hi-ris hi-fall mid-ris mid-fall lo-mid-lev lo-ris	

IDENTIFICATION DE LA LANGUE		
Nom :	Hausa	
Lieu :	N Nigeria (Kano)	
Population :	20 millions	(12 millions comme L1)
Classification :	Afro-Asiatic: Chadic: West: Group A: Hausa	
Consonnes	Occlusives: b b̄ t d d̄ k k ^w k ^j k ^ʔ k ^{ʔw} k ^{ʔj} g g ^w g ^j ʔ Affriquées: č ɲ Fricatives: φ s s ^ʔ z š h Nasales: m n Laterales: l Vibrantes: r ɾ Glides: j ʔ ^j w	
Voyelles	Antérieures: i e Centrales: ə Postérieures: u o Longues: ī ē ā ū ō Diphthongues: ai au	
Accent	Non pertinent	
Ton	haut, bas, descendant, downdrift	

La nomenclature employée a respecté autant que possible la nomenclature en vigueur, mais pour certaines rubriques comme le nom de certains éléments, le choix a été fait de les conserver en anglais, par exemple glide a été préféré à semi-consonne/semi-voyelle. Dans certaines situations, la nomenclature employée par les différentes sources a été reprise pour rester proche des descriptions proposées de ces langues.

Durant la phase de collecte des données, les langues ont été regroupées par continent (Afrique, Asie, Europe, Amérique...), ainsi pour chaque continent un fichier Excel relevant les différentes langues ainsi que l'existence ou non de tel ou tel élément a été proposé par la réalisation d'une matrice. L'une des possibilités qu'offrent la matrice que l'on a rempli avec des '+' et des '-' est que l'on peut aisément pour chaque continent indiquer le nombre de langues ayant tel ou tel élément par rapport au nombre total de langues de l'ensemble de la matrice. Un tel travail permet de fait de donner des informations relatives à tel marqueur dans une certaine zone géographique. A la fin du travail, nous avons opté pour remplacer les '+' par '1' et les '-' par '0', ce qui a permis d'arriver à proposer des sommes sur l'axe des horizontales (le nombre de consonnes de la langue) et sur l'axe verticale (le nombre exact de langues ayant par exemple la consonne p)

Pour proposer un aperçu global de l'ensemble des données recueillies, nous avons ensuite décidé de proposer un regroupement de l'ensemble des langues objet de la collecte de données dans un même fichier toujours sous Excel d'avoir une vue globale sur le corpus recueilli.

Les langues du continent Asie (1203 langues informées)

Language	Dialect	Classification	Country	Area	Continent	date name	Population	GLD number	seignement	β	η
1 Classical Mongolian		Altaic: Mongolian	Mongolia		Asia	(1) c. 1600 AD	1965	1	0	0	
2 Moghol	herat	Altaic: Mongolian: Western	Afghanistan	Herat	Asia	Moghol	5000	1966	1	0	
4 Dagur		Altaic: Mongolian: Eastern: Dagur	NE China		Asia	Daur	50000	1967	n.r.	0	
5 Monguor		Altaic: Mongolian: Eastern: Monguor	China	W Kansu	Asia	igghuer, Tu	60000	1968	1	0	
6 Yellow Uighur		Altaic: Mongolian: Eastern: Monguor	China	NW Kansu	Asia	West Yugur	4600	1969	n.r.	0	
7 Pao'an	Tung-hsiang	Altaic: Mongolian: Eastern: Monguor	China	Kansu	Asia	Baonan 10 (160000)		1970	1	0	
8 Santa		Altaic: Mongolian: Eastern: Monguor	China	SW Kansu	Asia		374000	1971	n.r.	0	
9 Oirat		Altaic: Mongolian: Eastern: Oirat-Khalkha: Oirat-Kalmuk	China		Asia		518000	1972	n.r.	0	
10 Kalmuk		Altaic: Mongolian: Eastern: Oirat-Khalkha: Oirat-Kalmuk	SW Russia	Kalmuk ASSR	Asia		253000	1973	1	0	
11 Buriat		Altaic: Mongolian: Eastern: Oirat-Khalkha: Khalkha-Buriat	SE Russia	Buriat ASSR	Asia		350000	1974	1	0	
12 Khalkha		Altaic: Mongolian: Eastern: Oirat-Khalkha: Khalkha-Buriat	Mongolia		Asia		1 million	1975	1	1	
13 Khorchin		Altaic: Mongolian: Eastern: Oirat-Khalkha: Khalkha-Buriat	China		Asia		900000	1976	1	0	
14 Ordos		Altaic: Mongolian: Eastern: Oirat-Khalkha: Khalkha-Buriat	China	Sui Yang	Asia		375000	1977	1	0	
15 Even		Altaic: Tungus: Northern	NE Russia	NE Yakut ASSR	Asia		9000	1978	1	0	
16 Negidal		Altaic: Tungus: Northern	SE Russia	Xabarovskij Krai	Asia		350	1979	1	0	
17 Evenki		Altaic: Tungus: Northern	E Russia	Yakut ASSR	Asia	Khamnigan	32000	1980	1	1	
18 Solon		Altaic: Tungus: Northern	E Russia	E Siberia	Asia			1981	n.r.	0	
19 Managir		Altaic: Tungus: Northern	E Russia	E Siberia	Asia			1982	n.r.	0	
20 Orochon		Altaic: Tungus: Northern	E Russia	E Siberia	Asia			1983	n.r.	0	
21 Ju'och'en		Altaic: Tungus: Southern: Southernwestern	E Russia	E Siberia	Asia		Extinct	1984	n.r.	0	
22 Manchu		Altaic: Tungus: Southern: Southernwestern	NW China	Sinkiang	Asia		27000 (50)	1985	1	0	
23 Manchu	Sibe	Altaic: Tungus: Southern: Southernwestern	China	Xinjiang	Asia	Xibe 100 (27000)		1986	1	0	
24 Akani		Altaic: Tungus: Southern: Southeastern: Nanaj	E Russia	E Siberia	Asia			1987	n.r.	0	
25 Birar		Altaic: Tungus: Southern: Southeastern: Nanaj	E Russia	E Siberia	Asia			1988	n.r.	0	
26 Gold		Altaic: Tungus: Southern: Southeastern: Nanaj	SE Russia	Xabarovskij Krai	Asia		7000	1989	1	0	
27 Kile		Altaic: Tungus: Southern: Southeastern: Nanaj	E Russia	E Siberia	Asia			1990	n.r.	0	
28 Olcha		Altaic: Tungus: Southern: Southeastern: Nanaj	SE Russia	Xabarovskij Krai	Asia		2000	1991	1	0	
29 Ork		Altaic: Tungus: Southern: Southeastern: Nanaj	SE Russia	Saxalin Is.	Asia		400	1992	1	0	
30 Samagir		Altaic: Tungus: Southern: Southeastern: Nanaj	E Russia	E Siberia	Asia			1993	n.r.	0	
31 Oroch		Altaic: Tungus: Southern: Southeastern: Udihe	SE Russia	Xabarovskij Krai	Asia		2000	1994	1	0	
32 Udihe		Altaic: Tungus: Southern: Southeastern: Udihe	SE Russia	Xabarovskij Krai	Asia		1400	1995	1	0	

Les langues du continent Amérique (1028 langues informées)

Colonne1	Colonne2	Colonne3	Colonne4	Colonne5	Colonne6	Colonne7	Colonne8	Colonne9	Colonne10	Co	Co	Co	Co	Co	Co	Co	Co	Co	Co	Co	Co	
998	Kaingang	Amerind: Macro-Ge: Kaingang	SE Brazil	Parana	South America		7000	5633	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
997	Xokleng	Amerind: Macro-Ge: Kaingang: Northern			South America			5634	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
998	Chiquil	Amerind: Macro-Ge: Kaingang: Northern			South America			5635	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
999	Amho	Amerind: Macro-Ge: Kaingang: Southern			South America			5636	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Ingain	Amerind: Macro-Ge: Kaingang: Southern			South America			5637	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Gualachi	Amerind: Macro-Ge: Kaingang: Southern			South America			5638	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1002	Acroa	Amerind: Macro-Ge: Ge: Cen	E Brazil		South America			5639	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Aricobe	Amerind: Macro-Ge: Ge: Cen	E Brazil		South America			5640	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Guegue	Amerind: Macro-Ge: Ge: Cen	E Brazil		South America			5641	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Chavante	Amerind: Macro-Ge: Ge: Cen	WC Brazil		South America		2000	5642	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Cherente	Amerind: Macro-Ge: Ge: Cen	E Brazil		South America			5643	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Aplisyne	Amerind: Macro-Ge: Ge: Nort	NC Brazil	Goiás	South America		200	5644	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1000	Suya	Amerind: Macro-Ge: Ge: Nort	E Brazil		South America			5645	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1000	Kreen-Akarore	Amerind: Macro-Ge: Ge: Northwest			South America			5646	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1010	Cayapo	Caraho	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5647	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	North	Amerind: Macro-Ge: Ge: Nort	C Brazil	S Para	South America	Kayapo	10000	5648	1	1	1	1	1	1	1	1	1	1	1	1	1
1011	Cayapo	Cruatire	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5649	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Curupite	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5650	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Duladi	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5651	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Gorotire	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America		250	5652	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Gradaho	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5653	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Pau d'Arco	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5654	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Cayapo	Uchicrin	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5655	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1011	Canela	Apanheca	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5656	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Augutje	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America		2000	5657	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Crabo, Krab	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5658	n.r.	1	1	1	1	1	1	1	1	1	1	1	1	1
1020	Canela	Crapimcaia	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5659	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Creye	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5660	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Cricatage	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5661	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Gurupy	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5662	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Macamecran	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5663	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Nucoecamed	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5664	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0
1020	Canela	Piocobge	Amerind: Macro-Ge: Ge: Nort	E Brazil	South America			5665	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0

Les langues du continent Océanie (1478 langues informées)

F1458	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1439	Mangata	Australian: Pama-Nyungan:		Australia	Oceania			4598	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1440	Nyangumarta	Australian: Pama-Nyungan:		NW Australia / Western Australia	Oceania		700	4599	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1441	Ngarla	Australian: Pama-Nyungan:		W Australia / Western Australia	Oceania		6	4600	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1442	Nyamal	Australian: Pama-Nyungan:		W Australia / Western Australia	Oceania		20	4601	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1443	Nyiyabali	Australian: Pama-Nyungan:		W Australia / Western Australia	Oceania	Balyku	50	4602	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1444	Tjurruru	Australian: Pama-Nyungan:		Australia	Oceania			4603	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1444	Kariyarra	Australian: Pama-Nyungan:		W Australia / Western Australia	Oceania		40	4604	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1448	Martuyunhira	Australian: Pama-Nyungan:		Australia	Oceania			4605	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1447	Nhuwala	Australian: Pama-Nyungan:		W Australia / W Western Australia	Oceania	Extinct		4606	n.r.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1448	Yindjibarndi	Australian: Pama-Nyungan:		W Australia / Western Australia	Oceania		200	4607	1	1	1																		

3. Exploitation possible à partir de la base de données aréales

- Création de matrice
- Comparaison des traits
- Comparaison Aréale
- Comparaison typologique

Une telle exploitation, comme le montrent l'exemple ci-dessous est un cas de comparaison typologique qui compare 12 langues selon la nature de leur système consonantique et de leur système vocalique. Ainsi, selon le type de voyelles ou de consonnes existant dans l'une ou l'autre langue, il sera possible de dégager des patrons phonologiques relatives à l'aire géographique où ces langues sont parlées.

Langue 1 : Manchu

Stops: p p^h t t^h k k^h q
 Affricates: č č^h qχ
 Fricatives: f v s (š) x χ
 Nasals: m n ŋ N
 Laterals: l
 Vibrants: r
 Glides: j
 Front Vowels: i y ε
 Central Vowels: ə a
 Back Vowels: u o
 Long Vowels: v̄
 Diphthongs: ai au ia ua əi əu iε yε oi io
 Stress: penult

Langue 2 : Korean

Stops: p p^h p^ʷ t t^h t^ʰ k k^h k^ʷ ʔ
 Affricates: č č^h č^ʷ
 Fricatives: s s^ʷ h
 Nasals: m n ŋ
 Laterals: l
 Glides: j w
 Front Vowels: i (y) e ø æ
 Central Vowels: a
 Back Vowels: u ɯ o ɤ
 Long Vowels: (v̄)

Langue 3 : Japanese

Stops: b t̚ d̚ k̚ g̚ (?)
 Affricates: č ʃ
 Fricatives: s z š h
 Nasals: m n̚ (ŋ)
 Vibrants: r
 Modified Consonants: c̄ p̄ t̄ k̄ č̄ s̄ š̄
 Glides: j w
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: ɯ o
 Long Vowels: v̄ all
 Tones: pitch accent: hi lo

Langue 4 : Garo

Stops: p^h b t̚ d̚ k̚ g̚
 Affricates: t̚s̚ d̚z̚
 Fricatives: s h
 Nasals: m n̚ ŋ
 Vibrants: r
 Glides: w
 Front Vowels: i e

Central Vowels: a

Back Vowels: u o

Langue 5 : Kapingamarangi

Stops: p p^h t t^h k k^h
 Fricatives: h
 Nasals: m m̚ n̚ ŋ̚ ŋ̚
 Vibrants: ɹ ʝ
 Glides: w w̚
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: u o
 Long Vowels: v̄ all
 Diphthongs: diphthongs
 Stress: penult or on long vowel

Langue 6 : Rapanui

Stops: p t k ʔ
 Fricatives: v h
 Nasals: m n ŋ
 Vibrants: r
 Glides: j w
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: u o
 Long Vowels: v̄ all
 Stress: penult

Langue 7 : Rarotongan

Stops: p t k ʔ
 Fricatives: β
 Nasals: m n ŋ
 Vibrants: r
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: u o
 Long Vowels: v̄ all

Langue 8 : Waskia

Stops: p b t d k g
 Fricatives: s
 Nasals: m n ŋ
 Laterals: l
 Vibrants: r
 Glides: j w
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: u o
 Stress: final

Langue 9 : Iwaidja

Stops: p t t̃ c k
 Fricatives: ɣ
 Nasals: m n ɲ ñ ŋ
 Laterals: l λ
 Vibrants: r r̃ ɹ
 Glides: j w
 Front Vowels: i e
 Central Vowels: a
 Back Vowels: u o
 Syllable: (C)V(C)(C)
 Stress: initial

Langue 10 : Gunwinygu

Stops: b d̃ d̃ ɟ ɟ̃
 Nasals: m ɲ n ñ ŋ
 Laterals: l l̃ λ
 Vibrants: r ɹ
 Glides: j w
 Front Vowels: i ε
 Central Vowels: a
 Back Vowels: u ɔ
 Long Vowels: v̄

Stress: non-phonemic

Langue 11 : Burarra

Stops: p t c k
 Nasals: m n ñ ŋ
 Laterals: l
 Vibrants: r ɹ
 Modified Consonants: c̃ p c k n ñ j
 Glides: j w
 Front Vowels: i æ
 Central Vowels: a
 Back Vowels: u ɔ

Langue 12 : Alawa

Stops: p^{mb} tnd tnd d̃ c̃ ŋ k ŋg
 Nasals: m n ɲ ñ ŋ
 Laterals: l l̃ λ
 Vibrants: r ɹ
 Modified Consonants: c̃ c n ŋ l
 Glides: j w
 Front Vowels: i ε
 Central Vowels: a
 Back Vowels: u ɔ
 Accent : penult

Langue	1	2	3	4	5	6	7	8	9	10	11
o	+	+	+	+	+	+	+	+	+	+	+
a	+	+	+	-	-	+	-	+	-	-	-
i	+	+	+	+	+	+	-	+	+	+	+
s	+	+	+	+	+	+	+	-	+	+	+
l	+	+	+	+	+	+	+	+	+	-	+
r	-	+	+	+	+	+	+	-	+	+	+
g	+	-	+	+	+	+	+	+	-	+	+
c	+	+	-	-	-	-	-	-	-	-	-
n	+	+	+	+	+	+	+	+	+	+	+
s	+	+	+	+	+	+	+	+	+	+	+
θ	+	+	+	+	+	+	+	+	+	+	+
o	+	-	+	-	-	-	-	+	-	+	-
u	-	+	-	+	+	-	-	-	-	+	+
p	-	+	+	-	+	+	-	-	-	-	-
u	-	-	+	-	-	+	-	-	-	+	+

Résultat de l'exploitation sous forme de matrice de comparaison selon la nature (existence ou non) de tel ou tel son consonantique ou de tel ou tel son vocalique

Une exploitation possible de ces données est envisagée dans le cadre de la reconstruction et de la recherche des protolangues. En effet, en partant des données fournies, il est tout à fait possible de créer des bases de données terminologiques et ainsi comparer des langues connues et considérées comme étant des langues de la même famille. En effet, les travaux en lexicostatistique peuvent se fonder sur une base de données de ce type pour construire un corpus et ensuite comparer le vocabulaire de base des langues en question pour mettre en lumière des protoformes et confirmer avec des données la parenté des langues mais aussi poser des hypothèses quant à la forme et à la structure es protoformes supposées. En tout état de cause, un tel travail s'inscrit de manière évidente dans le cadre de la typologie

des langues. Il faut noter que la question de la typologie en linguistique impliquent à la fois la classification des langues et la différenciation des langues au sein de groupes linguistiques spécifiques. Les approches synchroniques et diachroniques utilisent des typologies. Comment les langues sont-elles divisées en familles de langues et sous-groupes de langues ? Est-ce en fonction de critères de syntaxe, de vocabulaire, de morphologie et de phonétique ? Que nous apprend la répartition de ces familles linguistiques sur l'histoire des langues ?

En outre, la typologie en linguistique vise à classer les langues en familles et sous-familles linguistiques, tout en les distinguant également au sein d'un groupe linguistique donné, jusqu'à la granularité du dialecte ou de la variété. Les études typologiques ouvrent la voie à la reconstruction des langues originales, à l'étymologie et à la compréhension de l'Antiquité ou des migrations historiques, transcendant les oppositions traditionnelles entre langues flexionnes, agglutinantes ou isolantes, les catégories pouvant ne pas permettre de rendre compte des réalités propres aux diverses langues du globe. Désormais, sans négliger la grammaire, la comparaison des lexiques est privilégiée. La comparaison n'est pas l'apanage des linguistes, puisque tout apprenant d'une langue étrangère découvrira empiriquement la transparence, la superposition, les faux amis et la créativité sans limites qui se reflètent dans toutes les langues humaines. Ainsi, la base de données ainsi créée est un moyen de proposer un outil de comparaison entre les langues. On considère généralement que la typologie linguistique vise à établir des généralisations valables pour toutes les langues, indépendamment de leur affiliation à une famille de langues. Dès lors, les objectifs de la typologie linguistique paraissent éloignés de ceux de la linguistique indo-européenne, dont le but est au contraire de reconstruire un ancêtre commun par la comparaison de langues appartenant à une même famille de langues. En réalité, les liens entre linguistique indo-européenne et typologie linguistique sont de plus en plus affirmés depuis quelques années, à la fois par l'éclairage qu'apporte la typologie linguistique à la compréhension des faits propres aux langues indo-européennes et, en sens inverse, par l'apport de ces langues à la typologie linguistique.

Conclusion

Créer et structurer une base de données sur un certain nombre de langues du globe est une entreprise intéressante à plus d'un titre comme l'a montré les réalisations et exploitation possible notamment dans les travaux et recherches en typologie des langues. Réunir des données sur plus ou moins 5000 langues s'est fait en employant une méthode de recherche et de collectes propre à la typologie et se fondant sur les travaux de description et les grammaires existants. Nous posons que les exploitations illustrées ici ne sont pas exhaustives et il est donc possible d'envisager d'autres exploitations et d'autres développements à partir de cette base données.

Références bibliographiques

Chomsky, Noam, 1969, *Structures syntaxiques*, Paris, Seuil.

Culioli, Antoine, 1999, *Pour une linguistique de l'énonciation Tome 2 Formalisation et opérations de repérage*, Gap Paris, Éd. Ophrys.

Culioli, Antoine, 1999, *Pour une linguistique de l'énonciation Tome 3 Domaine notionnel*, Gap Paris, Éd. Ophrys.

De Saussure, Ferdinand, 2002, *Écrits de linguistique générale*, établis et édités par Simon Bouquet et Rudolf Engler (avec la collaboration d'Antoinette Weil), Paris, Gallimard.

Greenberg, Joseph, 1963, *The Languages of Africa*, Bloomington, Indiana University Press;

Meillet, Antoine, 1912, *Introduction à l'étude comparative des langues indo-européennes*, Paris, Hachette.

