

## LANGUES ET FAMILLES DE LANGUES (SCL 141-2002)

### Chap. I. Les langues en France:

Il ne faut pas croire qu'une langue n'est liée qu'à un seul pays, ni qu'un pays est un endroit où l'on ne parle qu'une seule langue. Le français par exemple n'est pas la langue qu'on parle en France, parce que d'une part le français est parlé dans d'autres pays que la France comme la Belgique, la Suisse, le Luxembourg, le Canada, et un certain nombre de pays d'Afrique, et parce que d'autre part on parle beaucoup d'autres langues en France que le français, même si le français est la seule langue officielle et nationale de la France.

En France, on parle en effet des langues qui appartiennent à des familles linguistiques parfois proches, mais aussi parfois très différentes.

**I. Les parlers romans**, c'est-à-dire issus du latin, sont bien représentés en France.

**A. Les 3 principales familles** de parlers romans:

**1. Les parlers occitans:**

"Les dialectes provençaux, ou occitans, ou d'oc couvrent le Midi de la France, ils sont limités au nord par une ligne est-ouest qui part de l'embouchure de la Gironde, coupe le département de la Dordogne à la hauteur de Périgueux, suit la limite septentrionale de la Haute-Vienne, la Creuse, le Puy-de-Dôme, s'infléchit vers le sud jusqu'à la hauteur de Valence et continue dans la direction de Turin à la limite de l'Isère et des Hautes-Alpes" (Pierre Guiraud, 1968, *Patois et dialectes français*, Que sais-je? n° 1285,29-30)

Quelques caractéristiques: "—absence de diphtongaison des voyelles accentuées libres. Aux français *miel*, *foi*, *douleur*, etc., correspondent *mel*, *fe*, *doulour*, etc. A ce même phénomène correspond la conservation de *a* qui en français s'est diphtongué en *ae* réduit à *é*. D'où les occitans *can* < prov. *chin* et *can* ), *prat*, *pan* en face de *chien*, *pré*, *pain*, etc.;

—conservation des consonnes occlusives intervocaliques qui sont tombées en français. D'où *vida*, *pagar*, *cadena*, <prov. *vido*, *paga*, *cadeno* > etc., en face de *vie*, *payer*, *chaene* (chaîne) <lat. *uita*, *pacare* «pacifier» —lat. pop. «apaiser, satisfaire» —a. fr. «payer» sens qui "s'est vraisemblablement développé sous l'influence du droit germanique qui, comme on sait, admettait le versement d'une somme d'argent en «payement» d'un dommage" (Bloch-Wartburg 447), *catena* >" (Pierre Guiraud, 1968, *Patois et dialectes français*, Que sais-je? n° 1285, 30).

**a. Notion de dialecte:** On qualifie ordinairement de dialectes les parlers occitans. Que faut-il entendre par là?

Les dialectes sont des variétés linguistiques qui ont la même origine, et qui reposent, chacune, "sur une certaine cohérence linguistique, mais aussi géographique, historique, ou plus largement culturelle" (Bouvier, Jean-Claude, 1979, "L'occitan en

Provence, Le dialecte provençal, ses limites et ses variétés", in: *Revue de linguistique romane*, 48-49). Comme le précise bien Jean-Claude Bouvier,

"la variété linguistique appelée dialecte <...> implique dans la vie de tous les jours l'existence ancienne ou la permanence d'un certain nombre de liens tissés par la volonté politique peut-être (Comté de Provence puis province de Provence, duché de Normandie...) mais surtout par l'habitude d'échanges de toute sorte, économiques, commerciaux, culturels, qui ont pu créer une certaine forme de conscience régionale" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 49). (cf. [transparent](#))

Concrètement, c'est par l'appartenance à un certain nombre d'aires dialectales que les dialectologues définissent au point de vue linguistique un dialecte. Les dialectologues appellent en effet *isoglosse* la ligne idéale qui, sur une carte, sépare deux traitements linguistiques différents, et *aire dialectale*, l'espace qui est à l'intérieur d'un isoglosse et qui par conséquent présente un même traitement linguistique.

Par exemple, sur la carte de la France, l'isoglosse de *a*, qui sépare le *a* latin accentué qui s'est diphtongué de celui qui s'est maintenu, ou, autre façon de présenter les choses, qui sépare une prononciation *pré* d'une prononciation *pra* (cf. [transparent](#)). Cet isoglosse sépare les langues d'oïl, qui sont au nord, de l'ensemble des langues d'oc et du franco-provençal, qui se trouve au sud. Mais cet isoglosse n'est pas exactement le même que celui de *a* après un *k*- initial comme le montre la carte de la *chèvre* (cf. [transparent](#)). De fait "en franco-provençal l'*a* se diphtongue derrière *k*- ce qui oppose l'occitan *cabro*, *tchabra* au franco-provençal *tchevra*" (Guiraud, Pierre, 1978, 3ème éd. (1ère éd.: 1968), *Patois et dialectes français*, Que sais-je? n° 1285, 31). Ainsi cet isoglosse ne réunit plus dans une même aire le domaine occitan et le domaine franco-provençal, à la différence de l'isoglosse *a/é* non précédé d'une dorsale de *pra/pré*.

Autre exemple: l'isoglosse *k*, qui sépare la prononciation palatalisée de *k*- initial latin de celle où il n'y a pas eu palatalisation, ou qui sépare *tchante* (ou *chante*) de *canta*. Entre ces deux parties d'isoglosse, on a l'aire dialectale de la palatalisation de *k*, laquelle couvre tout le domaine d'oïl moins le normand et le picard, le domaine franco-provençal et aussi toute la partie nord du domaine d'oc.

On voit bien ainsi que des isoglosses différents délimitent des aires dialectales différentes, et qu'un même domaine dialectal peut correspondre à des aires dialectales partiellement différentes, et se définit donc seulement par un certain faisceau de particularités qui n'ont pas toutes forcément —et loin de là— le même isoglosse. Si l'isoglosse de *a* diphtongué ou non diphtongué identifie bien les parlers d'oïl, il ne caractérise pas pour autant les seuls parlers d'oc, puisqu'il réunit les parlers d'oc et le franco-provençal. De même l'isoglosse de *d* intervocalique ou amuï n'est que partiellement pertinent pour délimiter les parlers d'oc, puisqu'il réunit le franco-provençal et une partie des parlers d'oc du nord-est. La

prise en compte des différences d'aire dialectale amène ainsi à postuler des subdivisions géographiques entre les dialectes d'un même domaine dialectal, ainsi que des chevauchements entre des domaines dialectaux différents.

**b. les dialectes provençaux** sont parmi les parlers occitans ceux qui sont parlés à l'Est du Rhône. "Comme l'indique la carte <cf. [transparent](#)>, l'espace dialectal provençal est traversé par deux grands axes qui se recoupent: l'un (A-B) horizontal allant d'Est en Ouest, l'autre (C-D) vertical, si l'on peut dire, allant approximativement du Nord au Sud en faisant une large boucle" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 56).

- **axe Est-Ouest**: "Le premier est bien connu puisqu'il parcourt toute la langue d'oc et distingue les parlers nord-occitans des parlers sud-occitans. Il sépare donc, dans notre région, nord-provençal et sud-provençal" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 56).

Au point de vue géographique, "la ligne de partage entre ces deux zones coïncide à peu près avec la limite départementale entre la Drôme et le Vaucluse, se confond avec le Ventoux et la montagne de Lure, passe ensuite au Sud de Digne, vers les Mées, Mezel, Barrême et remonte un peu vers le Nord dans les Alpes-Maritimes où elle passe entre Isola et Saint-Sauveur de Tinée" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 57).

Cette ligne de séparation correspond à l'isoglosse de *d* intervocalique ou amui (au Nord, on dit *maur* (lat. *maturu*), et au Sud *madur* «mûr». Autre différence: au Sud: *kabro* «chèvre», au Nord: *tsabro*, avec palatalisation de *k-* initial devant la voyelle *a*.

- **axe Nord-Sud**: "Le deuxième axe <...> permet de distinguer parlers alpins et non-alpins. Les parlers alpins situés dans les Hautes-Alpes, le Nord-Est des Alpes de Haute-Provence, une petite frange drômoise, les Alpes-Maritimes <...> ont <...> les principaux traits spécifiques suivants:

- conservation de *p*, *t*, *k*, *ts* en position finale.

- tendance au maintien de *-r* final dans les infinitifs.

- passage assez général de *-l-* intervocalique à *r* : *surèu* «soleil», *aro* «aile».

- diphthongaison de *è* en *èa* devant *r* : *vèart* «vert»." (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 57-58).

- **les 4 dialectes provençaux**: "Ainsi se trouvent délimités quatre sous-dialectes provençaux:

- 1) le «provençal alpin» qui possèdent à la fois les traits nord-provençaux et les traits alpins dont nous venons de parler <...>

- 2) le provençal «vivaro-dauphinois» qui <...> se situe dans une région non-alpine, qui a pour axe central le Rhône, la moitié orientale de l'Ardèche et la plus grande partie de la Drôme. Et il est bien moins conservateur que l'alpin: s'il assure la distinction entre singulier et

pluriel, ce n'est pas par le maintien d'-s au pluriel, mais par une opposition vocalique dans le féminin uniquement, sg. -o / pl. -a (comme en auvergnat et limousin) <...> (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 58)

3) le provençal «rhodano-méditerranéen» est le sous dialecte qui a de loin la plus grande extension géographique, puisqu'il comprend le Vaucluse, les Bouches-du-Rhône, le Var, une petite partie des Alpes de Haute-Provence et des Alpes-Maritimes. Ce qui unit tous ces parlers, c'est sans aucun doute, que <...> ils ont depuis le Moyen-Age connu des évolutions importantes dues principalement au contact avec le français: chute des occlusives finales et de s final, déjà signalée, et notamment abandon de la distinction entre singulier et pluriel dans les adjectifs et les substantifs: *l òme — liz òme; la fèmo — li fèmo*. (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 58-59).

"On peut bien sûr subdiviser ce groupe en «rhodanien» et «méditerranéen», conformément à l'usage. Importante du point de vue de l'histoire littéraire, puisque le rhodanien est la base du provençal mistralien, cette distinction est moins pertinente du point de vue scientifique: les traits qui peuvent départager ces deux variétés sont relativement nombreux mais souvent peu significatifs en raison d'interférences fréquentes." (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 59).

4) le «nissart» est peut-être le moins homogène des quatre sous-dialectes pour deux raisons. La partie la plus orientale des Alpes-Maritimes doit d'abord être détachée du nissart et rejoindre le grand ensemble dialectal de l'Italie de Nord ou plus précisément le dialecte ligurien" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 59).

"L'autre partie de cette originalité est constituée par le fait que, comme on le voit sur la carte, le nissart est au croisement des deux axes: la plupart des parlers nissarts sont tout à la fois sud-occitans et alpins" (Bouvier, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", 61).

**c. les autres dialectes occitans:** Les 4 dialectes provençaux ci-dessus mentionnés sont ordinairement désignés par les deux étiquettes générales de **provençal** pour les deux dialectes du Sud, et de **vivaro-alpin** pour les deux dialectes du Nord, la séparation entre ces deux zones dialectales étant ce que nous avons appelé l'axe est-ouest. Cet axe est-ouest se prolonge dans le reste du domaine occitan à l'ouest du Rhône, séparant ainsi les parlers nord-occitans des parlers sud-occitans.

A l'ouest du Rhône, les parlers nord-occitans sont formés principalement du **limousin** et de l'**auvergnat**, et les parlers sud-occitans de ce qu'on appelle le **languedocien**. A l'intérieur du languedocien, on peut distinguer "un *languedocien méridional* central, toulousain, fuxéen, donésanais, narbonnais), qui se rattache à l'ensemble aquitano-pyrénéen <...>, un *languedocien septentrional*, qui fait la transition avec les parlers nord-occitans (rouergat, gévaudanais, aurillacois), un *languedocien occidental* (agenais, quercinois,

albigeois), et un *languedocien oriental*, qui annonce déjà le provençal (biterrois, montpelliérain, cévenol)" (Bec, Pierre, 1978, 4ème éd. (1ère éd.: 1963), *La langue occitane*, Que sais-je? n° 1059, 44-45).

Toutes ces informations relèvent plus de l'histoire que de l'actualité. Car, comme le dit Georges Kersaudy, "l'occitan est incontestablement la plus menacée de nos langues minoritaires. <...> Les parlers occitans, éparpillés sur une aire géographique très vaste, sont particulièrement vulnérables, en raison de l'uniformisation imposée par les media et par l'enseignement officiel. Il serait difficile de dire combien de méridionaux en France utilisent l'occitan sous une forme quelconque et dans quelles circonstances, car on ne dispose à cet égard d'aucune donnée statistique; ils sont probablement des millions. De nos jours, même dans les grandes villes, certains essaient de maintenir l'usage de l'occitan, mais malgré les efforts remarquables déployés par des organismes bénévoles, la langue a beaucoup de mal à faire reconnaître son existence et à retenir l'attention des pouvoirs publics" (Kersaudy, 2011, 45-46).

Il faut signaler qu'une querelle divise les gens qui pratiquent ou enseignent la langue d'oc; c'est le problème de l'écriture ou de l'orthographe. Il y a en effet deux graphies différentes, qui, du moins en théorie, sont séparées par le Rhône, et que l'on appelle la graphie occitane et la graphie mistralienne. La première, qui se fonde sur l'usage des troubadours, a été redéfinie par l'Institut d'Etudes provençales de Toulouse, qui l'a même appliquée aux parlers de la Provence. Elle se veut unitaire cherchant à transcender les différences de prononciation entre les dialectes occitans. La seconde a été inventée au 19<sup>ème</sup> siècle par un groupe de 7 poètes provençaux, appelé le Félibrige, et animé par Mistral, prix Nobel de littérature en 1904. Elle entend noter la prononciation d'un dialecte choisi comme langue littéraire, à savoir le dialecte de Maillane (village des Bouches du Rhône, près de Saint Rémy de Provence) dans lequel Mistral a écrit son épopée *Mireille*. Jean-Claude Bouvier a bien essayé de proposer un compromis unitaire entre ces deux graphies, fondé sur les deux principes suivants:

"1) le besoin de rendre la graphie facile à lire par tout le monde, et 2) la nécessité de concilier l'unité et la diversité de la langue occitane" (d'après Bouvier, 1979 (?), "Noto sus la grafio", in: *Lo biais de se dire* (Vitrolles), 1, 3).

Mais cela n'a pas eu beaucoup de succès.

**d. Le gascon et le catalan** sont des parlers romans proches de l'occitan, mais qui néanmoins s'en distinguent.

- **le gascon**: au sud-ouest de la Garonne, se rattache aux parlers occitans, mais a une place à part: "les *f* du latin y sont prononcés avec un *h* (un vrai *h* expiré): *hiria* «farine» du latin *farina*), *hlor* «fleur» (du latin *flor*) ou *hilha* «fille» du latin *filia*)" (Henriette Walter, 1994, *L'aventure des langues en occident*, 249).

• **le catalan (*català*)**: "L'ensemble catalan se partage entre la France (Roussillon et Cerdagne) et l'Espagne" (Henriette Walter, 1994, *L'aventure des langues en occident*, 247). Le catalan ressemble au languedocien, dont il se distingue par plusieurs traits. On citera par exemple le maintien du [u] latin. "*MADURU(M)* a évolué en *madur* (prononcé *madour*) «mûr», *LUNA* <en> *lluna* <prononcé <prononcé [λuna]> «lune»" (Henriette Walter, 1994, 248), en face de pr. *madur* et *luno*.

"C'est en 1976 que la première école maternelle en catalan, la *Bressola*, a été créée à Perpignan et, depuis 1982, un cursus complet d'enseignement du catalan a été mis en place à l'université de Perpignan" (Henriette Walter, 1994, *L'aventure des langues en occident*, 247). "Parlé en France par 200.000 habitants dans le Département des Pyrénées orientales", le catalan, "langue majoritaire et officielle en Catalogne et en Andorre <est> la langue de dix millions d'Européens. Après de longues années de stagnation, <il> est aujourd'hui en plein essor, et c'est probablement la langue minoritaire <de France> pour laquelle les perspectives sont les plus brillantes dans le cadre de l'Union Européenne" (Kersaudy, 2001, 46).

## 2. Les parlers d'oïl, ou les parlers français:

"Voici les principales divisions dialectales du domaine français:

—le Centre: autour du *francien*, <prétendument> dialecte de l'Île-de-France et source du français moderne, groupe *l'orléanais*, *le bourbonnais*, *le champenois*;

—le Nord: *le picard* (Picardie, Artois), *le haut-normand*, *le wallon*; <ils se caractérisent par exemple par la conservation de *k* devant *a* (*quèvre*, *kèvre* au lieu de *chèvre*, cf. lat. *capra*) > (cf. [transparent](#) carte dialectale de *chèvre* )

—l'Est: *le lorrain*, *le bourguignon*, *le franc-comtois*;

—l'Ouest: *le bas-normand*, *gallo* (basse Bretagne), *l'angevin*, *le maine*;

—le Sud-Ouest: *le poitevin*, *saintongeais*, *angoumois* " (Pierre Guiraud, 1968, *Patois et dialectes français*, Que sais-je? n° 1285, 34-35).

Le francien est, d'après Bernard Cerquiglini, une pure invention. Les linguistes ont cru que le français littéraire national ne pouvait pas ne pas être le triomphe d'un dialecte particulier. Ils ont donc parlé du dialecte de l'Île-de-France. Mais les dialectologues sont bien incapables de préciser ce qui caractérisait ce dialecte. Ils se contentent de dire ce qu'il n'était pas, en affirmant qu'il n'avait pas les particularités des autres différents dialectes voisins.

"La genèse d'un usage écrit, traditionnel et interrégional, n'est <...> pas la promotion politique d'un dialecte particulier, le *francien*. Elle est une pratique, qui tend à constituer un *français*, langue des lettres et des lettrés. <...> elle peut certes se lier au politique (ce que les *Serments <de Strasbourg>* montrent clairement), mais elle est issue du milieu des clercs, désireux de fabriquer des textes en français, et souhaitant pour cela mettre au point une langue qui, avec la même dignité que le latin, voire une pérennité et une universalité comparables, puisse dire le courage du héros, l'amour de Dieu et des femmes" (Cerquiglini, 1993<sup>2</sup>, *La naissance du français*, Que sais-je? n° 2576, 118)

En réalité la survivance de ces dialectes est de plus en plus douteuse, surtout dans la partie centrale du domaine; car "au Centre une vaste zone, qui prend de plus en plus d'importance, ignore tout autre parler que le français" (Bernard Pottier, "La situation linguistique en France", in: *Le langage*, 1968, 1154). De fait "les formes des mots issus du latin y <furent> souvent très proches de celles qui ont été perpétuées en français. La perte de l'idiome local se <fit> alors, d'abord imperceptiblement, puis par fusion progressive avec le français, qui finit par l'absorber" (Henriette Walter, 1994, *L'aventure des langues en occident (Leur origine, leur histoire, leur géographie)*, Paris, Laffont, 249).

**3. le franco-provençal:** "Situé entre le provençal et le français dont il partage certains caractères, le franco-provençal part de Thiers à l'extrémité nord-est du département du Puy-de-Dôme et remonte jusqu'à Montbéliard et Bâle. Il comprend les départements de la Savoie, de l'Isère, de la Loire, du Rhône, de la Haute-Savoie, de l'Ain ainsi que le sud du Jura, le Doubs, la Suisse romande et le Val d'Aoste." (Pierre Guiraud, 1968, *Patois et dialectes français*, *Que sais-je?* n° 1285, 33-34).

Caractéristiques: "Le franco-provençal est essentiellement caractérisé par la conservation de *a* accentué libre, caractère qu'il partage avec le provençal; en revanche les occlusives intervocaliques qui résistent en provençal s'amouissent en franco-provençal comme en français. Ainsi le latin *prata* donne en <anc.> français *prée* <<prairie>>, en provençal *prado* <<prairie>> et en franco-provençal *praye*" (Pierre Guiraud, 1968, *Patois et dialectes français*, *Que sais-je?* n° 1285, 34).

fr. *cheval*, fr-pr. *chival*, prov. *cavau* (emprunt: *chivau* ).

**B. Les parlers corses** représentent un quatrième groupe de parlers romans; ils "font partie du groupe linguistique italien. Au nord de l'île, on parle le toscan; au sud également, mais avec des marques d'influence sarde" (Bernard Pottier, "La situation linguistique en France", 1157).

Le corse (*corsu*) est en effet très proche de l'italien. Une petite différence: "le corse a gardé certains traits anciens du latin disparus en italien, comme par exemple la voyelle *u* (prononcée *ou*) en position finale" (Henriette Walter, 1994, *L'aventure des langues en occident*, 247):

cors. *pràtu*, ital. *prato*, fr. *pré* (lat. *pratum*)

cors. *pèlu*, ital. *pelo*, fr. *poil* (lat. *pilus*)

cors. *chiàru*, ital. *chiaro*, fr. *clair* (lat. *clarus*)

cors. *dìtu*, ital. *dito*, fr. *doigt* (lat. *digitus*).

Il est "parlé par environ 200.000 personnes, tant en Corse que sur le continent". Il est "largement majoritaire sans l'île et sa position semble actuellement se renforcer" (Kersaudy, Georges, *Langues sans frontières. A la découverte des langues de l'Europe*, Editions Autrement, 2001, 45).

## II. Les parlers non romans:

### 1. Parlers indo-européens:

**a. celtiques:** "le breton (*brezhoneg*) fait partie du groupe «brittonique» des langues celtiques. On distingue quatre variétés régionales; le vannetais, le cornouaillais, le trégorrois et le léonais <=léonard>, ce dernier ayant une certaine suprématie, dans la langue écrite notamment" (Bernard Pottier, "La situation linguistique en France", 1158).

Répartition géographique:

- Le cornouaillais (*kerne*) est parlé dans le sud Finistère, où se trouve la préfecture Quimper, et dans une partie des deux départements voisins: les Côtes-d'Armor, au nord, et le Morbihan, au sud.

- Le léonard (*leon*) est parlé dans le nord Finistère, dont Brest et Morlaix sont les deux sous-préfectures.

- Le trégorrois (*treger*) est parlé au nord de la Bretagne, dans les Côtes-d'Armor, dont Saint-Brieuc est la préfecture, mais ne se trouve plus maintenant dans la zone où est pratiqué le trégorrois.

- Le vannetais enfin est parlé au sud, dans le Morbihan, dont Vannes est la préfecture, et est maintenant l'extrême limite-est du dialecte vannetais.

Il est certain que le breton a beaucoup régressé depuis le 9<sup>ème</sup> siècle, époque où toute la Bretagne, de la baie du Mont Saint Michel à l'estuaire de la Loire, parlait le breton. Au 20<sup>ème</sup> siècle, la limite orientale du breton va approximativement de Paimpol à Vannes. On peut ramener ces 4 dialectes bretons à 2 zones dialectales, en regroupant le cornouaillais (en breton *kerne*), le léonard et le trégorrois), ensemble dialectal que l'on désigne couramment sous le sigle KLT, et en opposant globalement ces 3 dialectes au seul vannetais. Deux particularités phonétiques au moins distinguent ces deux dialectes:

"— ce qui se prononce [z] en KLT se prononce [h] en vannetais (comme le *h* de l'anglais ou de l'allemand);

— l'accent se trouve sur l'avant-dernière syllabe en KLT, et sur la dernière en vannetais <...>. On pourrait encore ajouter que la grammaire et le vocabulaire du vannetais se différencient nettement de ceux des trois autres dialectes. Toutes ces raisons ont suggéré l'hypothèse que le vannetais pourrait bien être une survivance du gaulois continental, alors que les trois autres dialectes conservent des ressemblances plus nettes avec le celtique insulaire" ((Henriette Walter, 1994, *L'aventure des langues en occident*, 97).

Parmi les bretonnants, on peut distinguer actuellement deux groupes différents. "Le premier est un petit groupe de gens instruits —environ 20.000— <...> qui n'ont appris le breton qu'à l'adolescence et qui oeuvrent pour sa reconnaissance, son enseignement et sa diffusion. L'autre groupe est beaucoup plus nombreux —on avance le chiffre de 500.000— mais il décroît rapidement. Il est constitué par les couches les plus âgées des milieux ruraux, où les différences régionales sont plus accentuées" (Walter, 1994, 99-100). A l'intérieur du premier groupe, il y a deux écoles rivales, qui se distinguent notamment par une querelle

orthographique. Autour de l'Université de Brest, on prône "l'orthographe universitaire" proposée en 1956 par le chanoine François Falc'hun [fal un], et autour de l'Université de Rennes, on emploie le breton ou la graphie *zh*, c'est-à-dire une graphie unifiée. Celle-ci par exemple écrit *Breizh* le nom de la *Bretagne*, qui se prononce *Breiz* en Cornouaille, dans le Léon et le Trégorrois, mais *Breih* en vannetais.

#### **b. germaniques:**

- **le flamand (*vlaamsch*):** "Le flamand fait partie du groupe linguistique néerlandais (germanique occidental)", très précisément de la branche dite bas-allemand (*niederdeutsch*) du germanique occidental, encore appelé westique. "Il s'est sensiblement différencié du hollandais littéraire. Le flamand est parlé au nord d'une ligne passant à l'est de Dunkerque et Saint-Omer et au sud de Bailleul" (Bernard Pottier, "La situation linguistique en France", 1159).

"Dans la région de Dunkerque appelée «Westhoek», environ 200.000 Français parlent encore le dialecte flammand du pays, et la plupart d'entre eux, tout en pratiquant le français, sont capables de lire en néerlandais" (Kersaudy, 2001, 44).

- **Le francique lorrain:** "ce parler germanique — appelé *Lothringer Platt* — se rattache à ceux de l'Allemagne moyenne (Rhénanie, Sarre, Palatinat), de la Belgique (Arlon, Saint-Vith) et du grand-duché du Luxembourg. Il occupe en France plus de la moitié du département de Moselle et une partie au nord du Bas-Rhin" (Henriette Walter, 1994, *L'aventure des langues en occident*, 327). Le westique ou germanique occidental se subdivise en bas-allemand (*niederdeutsch*) et haut-allemand (*hochdeutsch*), la bas-allemand étant approximativement au nord, et le haut-allemand au sud d'une ligne Cologne-Berlin. Le haut-allemand se subdivise à son tour en allemand moyen (*mitteldeutsch*) et allemand supérieur (*oberdeutsch*), l'allemand supérieur étant l'allemand méridional, et l'allemand moyen se trouvant entre le bas-allemand et l'allemand supérieur. Le francique lorrain fait donc partie de l'allemand moyen, tandis que l'autre dialecte germanique parlé en France, qui est appelé alsacien, fait partie de l'allemand supérieur ou méridional. On sait que l'allemand langue commune provient de la traduction par Luther de la Bible, c'est-à-dire de la partie orientale de l'allemand moyen.

- **"L'alsacien** aujourd'hui se restreint de plus en plus à l'usage rural, familial et intime, mais il n'est pas rare à Strasbourg d'entendre les gens s'exprimer en alsacien dans les boutiques, à la poste ou au marché" (Henriette Walter, 1994, *L'aventure des langues en occident*, 327).

**2. Langue non indo-européenne:** "le basque (*euskara* ou *euskuara*) est une langue non indo-européenne. Le basque de France se divise en trois zones: le labourdin (ouest), le bas-navarrais (centre), le souletin (est). <...> Cette langue est parlée au sud-ouest d'une ligne partant du sud de Biarritz-Bayonne, et passant au nord de Saint-Palais, à l'est de

Mauléon et de Tardets; toutes ces localités sont situées dans le département des Basses-Pyrénées" (Bernard Pottier, "La situation linguistique en France", 1159).

On ne sait pas sûrement à quelles langues il faut apparenter le basque. Une hypothèse, qui ne fait pas l'unanimité, mais qui a pour elle notamment l'autorité de Dumézil, rattache le basque aux langues caucasiennes. Dumézil a écrit:

"Entre le caucasien du Nord (surtout du Nord-Ouest) et le basque, dans toutes les parties de la grammaire, des analogies précises de structure et de remarquables coïncidences de détail ont été signalées <...>. Entre le caucasien du Sud et le basque, peu d'analogies de structure, mais d'assez nombreuses rencontres de vocabulaire (au moins une centaine)" (Georges Dumézil, "Langues caucasiennes", in: Meillet et Cohen (éd.), 1952 (2ème éd.), *Les langues du monde*, Champion, 228).

Le basque "est parlé en France par environ 100.000 personnes <et> en Espagne <...> par près d'un million d'habitants" (Kersaudy, 2001, 45). Actuellement "une sorte de «basque unifié» (*Euskara batua*) tend à se dégager, qui prend pour base un compromis entre le navarro-labourdin, langue de référence habituelle dans le Pays basque français, et le guipuzcoan, qui est la forme de basque la plus volontiers prise pour norme en Espagne, bien que le biscayen ait le plus grand nombre de locuteurs. L'Académie de la langue basque, *Euskaltzaindia*, créée en 1919, en a fixé le système graphique, alors qu'au 17<sup>e</sup> siècle il suivait les règles du français en France et celles du castillan en Espagne" (Walter, 1994, 184).

## Chap. II. La famille des langues romanes

Nous avons vu que le français, langue nationale, fait partie des dialectes d'oïl, ou plus exactement a été formé à partir des dialectes d'oïl, dialectes qui couvrent la partie nord de la France. Et nous avons vu que d'autres dialectes couvrent la partie sud de la France, à savoir les dialectes d'oc, et qu'entre ceux deux zones dialectales, plus précisément au centre est de la France, il y a une 3<sup>ème</sup> zone dialectale appelée franco-provençal. Ces 3 groupes de dialectes appartiennent à ce qu'on appelle les langues romanes, c'est-à-dire l'ensemble des langues de l'Europe qui sont issues de la langue latine, et sont par conséquent apparentées. On dira donc que les langues romanes est une famille de langues, et même plus précisément une famille d'ensembles dialectaux, dans la mesure où une langue comme le français représente déjà un ensemble dialectal assez complexe.

**1. Similitudes entre langues romanes:** Si, à la façon des dialectologues, on compare les mots qui sont utilisés pour désigner le père, la mère et le frère dans les différentes langues parlées dans le sud de l'Europe, c'est-à-dire, en allant de gauche à droite sur la carte, depuis l'Espagne à l'Ouest jusqu'à la Roumanie à l'Est, en passant par le Portugal, la France, la Suisse, l'Italie, ainsi qu'à date ancienne la Yougoslavie, on constate qu'il y a à la fois de

grandes ressemblances et d'incontestables différences entre les termes utilisés, comme le montre bien le tableau suivant:

	<b>père</b>	<b>mère</b>	<b>frère</b>
Espagnol	<i>padre</i>	<i>madre</i>	hermano
Portugais	<i>pai</i>	<i>mãe</i>	irmão
Catalan	<i>pare</i>	<i>mare</i>	germà
Provençal	<i>paire</i>	<i>maire</i>	<i>fraire</i>
Français	<i>père</i>	<i>mère</i>	<i>frère</i>
Corse	babbu	mamma	fratellu
Italien	<i>padre</i>	<i>madre</i>	fratello
Sarde	babbu	mama	<i>frade</i>
Sicilien	<i>patri</i>	<i>matri</i>	<i>frati</i>
Roumain	tatfa	mamfa	<i>frate</i>

où sont en italiques les termes qui manifestement présentent des ressemblances, mais où les termes qui ne sont pas en italiques ne sont pas sans rapport entre eux. De fait, parmi les mots apparemment divergents, il y a des ressemblances entre les mots espagnol, portugais et catalan d'une part, corse et italien d'autre part qui correspondent à «frère», entre les mots corse, sarde et roumain qui correspondent à «mère», et entre les mots corse et sarde qui correspondent à «père».

Ces ressemblances ne sont pas dues au hasard, comme l'établit la linguistique comparée, qui va montrer qu'il s'agit de bien plus que de simples ressemblances. D'autre part elles ne surprennent pas l'historien qui sait qu'entre le 1er siècle avant J.-C. et le 5ème siècle après J.-C., cet espace géographique a été progressivement conquis et administré par les Romains.

Voici en effet un petit rappel historique. En 227 avant J.-C., après la 1ère guerre punique, les Romains font de la Sicile la première province romaine, et créent les provinces de Corse et de Sardaigne. En 197, l'Espagne devient province romaine, avec deux préteurs, l'un pour l'Espagne Ulérieure, l'autre pour l'Espagne Citérieure. En 120/119, les Romains créent la province de Gaule narbonnaise. En 51, après la guerre des Gaules, c'est la Gaule celtique qui devient province romaine. En 15 avant J.-C., la Rhétie (région alpine du Tyrol et de la Bavière) et le Norique (région entre la Rhétie et la Pannonie) deviennent provinces romaines. Et finalement, création, en 10 après J.-C., de la province de Pannonie (= Hongrie), et en 107, de celle de Dacie (= Roumanie).

On connaît donc la langue-mère qui explique ce qu'il y a de commun entre toutes les langues de cet espace géographique. Il s'agit du latin, langue des Romains. Voilà pourquoi

on qualifie de romanes ces différentes langues, qui sont manifestement issues du latin. Mais le latin en question n'est pas celui des auteurs latins classiques, que l'on étudie encore de nos jours, lequel est une langue littéraire et cultivée. Il s'agit du latin parlé tardif et populaire, que l'on appelle traditionnellement le latin vulgaire. On en a quelques traces dans tous les documents littéraires ou non qui furent plus ou moins fortement influencés par la langue parlée des gens non cultivés, comme les inscriptions de Pompéi, le *Satiricon* de Pétrone ou l'*Appendix Probi*, supplément à un traité grammatical du pseudo Probus qui est un recueil de fautes du type "Ne dites pas..., mais dites...".

De fait, les mots par lesquels le plus grand nombre des langues romanes désignent le père, la mère et le frère remontent à l'accusatif des mots qui signifiaient «père», «mère» et «frère» en latin, à savoir *pátre(m)*, *mátre(m)* et *frátre(m)*, mots qui figurent comme entrée dans les dictionnaires latins sous la forme des nominatifs singuliers: *pater*, *mater* et *frater*.

**2. Linguistique comparée:** La linguistique comparée ne se contente pas de vagues ressemblances entre les langues pour les rattacher à une même langue-mère. Elle se fonde sur ce qu'elle appelle des règles de correspondances, c'est-à-dire des formules qui établissent une correspondance systématique entre un son donné d'une langue et un autre son particulier, plus ou moins différent, de chacune des autres langues concernées; et ces règles de correspondances peuvent être illustrées par un certain nombre d'exemples.

**a. Quelques règles de correspondances:** C'est ainsi que les différents mots signifiant «père» supposent la règle de correspondances concernant [p] initial:

port. p- = esp. p- = cat. p- = prov. p- = fr. p- = it. p- = sc. p-

(le sarde, le corse et le roumain ne peuvent pas être pris en compte, puisqu'ils utilisent, pour la notion de «père» des mots très différents de ceux des autres langues romanes; mais d'autres séries de mots permettraient, comme nous le verrons par la suite, d'ajouter à cette formule: = sar. p- = cors. p- = roum. p-). Cette règle de correspondances est confirmée notamment par les séries de mots suivantes:

port. Prado, esp. Prado, cat. Prats, prov. Prats, fr. Pré, cors. Pràtu, it. Prato, ~~roum.~~ Pajiște (cf. v. sl. pašiste)

port. Ponte, esp. Puente, cat. Pont, prov. Pont, fr. Pont, cors. Ponte, it. Ponte, roum. Punte

port. Pêra, esp. Pera, cat. Pera, prov. Pero, fr. Poire, cors. Pèra, it. Pera, sc. Pira, sar. Pira, roum. Păra

port. Pêlo, esp. Pelo, cat. Pèl, prov. Péu, fr. Poil, cors. Pèlu, it. Pelo, sc. Pilu, sar. Pilu, roum. Păr

port. Pobre, esp. Pobre, cat. Pobre, prov. Paure, fr. Pauvre, cors. Pòveru, it. Povero, ~~roum.~~ Sărac (cf. bulg. sirak).

Il peut arriver que certaines langues utilisent un autre mot que celui qui devrait correspondre aux mots d'origine latine des autres langues, comme c'est le cas dans les exemples précédents

pour le roumain. Une telle absence de correspondance n'est nullement un contre-exemple à la règle de correspondances. Cela veut seulement dire que la langue concernée a fait un emprunt à une autre langue (comme une langue slave dans le cas du roumain), et que cet emprunt a supplanté le mot qui lui venait du latin.

Les deux mots qui signifient «père» et «mère» mettent en évidence une autre règle de correspondances concernant des consonnes intérieures, à savoir:

port. -dr- = esp. -dr- = fr. -r- = cors. -dr- = it. -dr- = roum. -tr-

que l'on peut confirmer par les exemples suivants:

esp. ladrón, port. ladrão, cat. lladre, fr. larron, cors. ladrone, it. ladrone, ~~roum.~~-hot§

port. quadro «1. carré, 2. cadre, 3. tableau», esp. cuadro, cat. quadró, fr. carré, cors. quàdru, it. quadro, roum. pătrat.

Certes, dans cette formule de correspondances, toutes les langues ont en commun une vibrante. Mais on notera que cette vibrante n'est pas phonétiquement identique dans toutes les langues comparées. Elle est en effet uvulaire en français et apicodentale ailleurs. Il y a donc bien correspondance entre ces différentes langues, et non pas identité phonétique. On voit également que le français n'a qu'un seul phonème là où les autres langues ont un groupe de phonèmes, et que l'occlusive apicodentale qui, dans les langues avec groupe de phonèmes, précède la vibrante est tantôt voisée tantôt sourde. Il y a donc à la fois identités et différences dans cette formule de correspondances. Et c'est la systématisme de telles correspondances et non la vague ressemblance ou l'identité accidentelle qui est la preuve de la parenté entre les langues.

Les trois mots signifiant «père», «mère» et «frère» mettent en évidence une troisième règle de correspondances qui concerne les voyelles. Il s'agit de la règle:

port. a = esp. a = cat. a = fr. è = cors. a = it. a = sc. a = sar. a = roum. a

ce qu'on peut confirmer par les séries de mots suivantes:

port. claro, esp. claro, cat. clar, fr. clair, cors. chiàru, it. chiaro, roum. clar

port. caro, esp. caro, cat. car, fr. cher, cors. caru, it. caro, ~~roum.~~ drag (cf. v.sl. dragŭ)

port. escala «escale», esp. escala «échelle», cat. escala, fr. échelle, cors. scàla, it. scala, roum. scară.

Ici le français fait correspondre une voyelle intermédiaire [ɛ] à la voyelle ouverte [a] des autres langues romanes.

Ce ne sont pas de simples ressemblances partielles qui permettent d'établir la parenté entre des langues, mais de véritables règles de correspondances qui font systématiquement correspondre des sons identiques ou non identiques entre lesdites langues. C'est la systématisme de la correspondance qui ne peut être due à un simple hasard va obligatoirement s'expliquer par un héritage commun. L'arbitraire du signe linguistique, c'est-à-dire le fait qu'il n'y ait aucun rapport nécessaire entre le signifiant et le signifié des signes linguistiques d'une langue exclut que des signifiés identiques ou proches présentent des

signifiants phonétiquement identiques, proches ou très différents qui soient systématiquement comparables dans des langues différentes. Seule une origine commune peut expliquer ce que le hasard ne saurait expliquer. On remarquera que les mots comparables se conforment en fait à plusieurs règles systématiques de correspondances, et même que chacun des phonèmes qui constituent leur signifiant applique une règle de correspondances, ce qui exclut encore plus la possibilité d'un simple hasard. Toutefois on observe que des langues, manifestement apparentées du fait des règles de correspondances comme les langues romanes, ne présentent pas que des lexèmes comparables. On a vu que des emprunts peuvent expliquer qu'une langue n'ait pas un lexème apparenté à ceux des autres langues: le roumain par exemple a emprunté des lexèmes aux langues qui l'entouraient comme le vieux slave ou le bulgare. Mais il arrive aussi que des divergences lexicales concernent non plus une seule langue, mais plusieurs des langues théoriquement apparentées, comme on peut le voir dans les trois séries de termes de parenté sur lesquelles nous travaillons.

**b. Les mots non comparables ou divergents:** Les trois séries de termes de parenté sont particulièrement intéressants; car, dans leur divergence ils établissent d'autres parallélismes et d'autres correspondances, qui, loin de contredire l'hypothèse de la parenté génétique entre les langues romanes, la renforcent et la confirment. Elles montrent en effet que, si le lexique n'a pas évolué de la même façon dans toutes les langues romanes, celles-ci n'ont pas forcément

Ainsi dans le cas de «frère», l'italien *fratello* et le corse *fratellu*, qui sont manifestement parallèles entre eux, ne sont nullement l'équivalent de:

roum. frate, sar. frade, sicil. frati.

Toutefois l'équivalent phonétique de ces derniers mots existe en italien et en corse, mais il a un autre sens que celui de «frère»; c'est:

cors. frate «moine», it. frate «moine» (cf. *frati minori* «frères mineurs» ou le titre de *fra* «frère» dans les couvents).

Les deux mots *fratello* et *fratellu* sont manifestement des dérivés diminutifs de ce *frate*. Ils devaient donc signifier à l'origine «petit frère»; et ils ont certainement pris le sens de «frère» tout court, pour remplacer *frate*, quand ce dernier mot s'est spécialisé pour signifier les «frères en religion», c'est-à-dire les «moines».

La série de mots signifiant «frère» contient d'autres mots qui n'ont rien à voir ni avec roum. *frate* ni avec ital. *fratello*, mais qui présentent des ressemblances que l'on peut définir très rigoureusement par d'autres règles de correspondances. Il s'agit de:

esp. hermano (a. esp. ermano), port. irmão, cat. germà

D'abord la finale des deux premiers mots repose sur la correspondance:

port. -ão = esp. -ano

que manifestent la série de mots suivante:

esp. grano, port. grão, fr. grain; esp. mano, port. mão, fr. main; esp. llano «plat», port. chão (< lat. planu) «terre».

Ensuite quand on sait que dans esp. *hermano*, "L'h initial est ici purement graphique", on a alors les correspondances suivantes:

esp. j- (mais Ø- devant voyelle atone) = port. j- [ʒ] (mais Ø- devant voyelle atone) = it. g- ([dʒ]) = cat. j-

que mettent en évidence les séries de mots suivantes:

esp. helar «geler» (a. esp. elar), it. gelare, cat. gelar;

esp. hinojos «les genoux» (a. esp. inojos), it. ginocchio «genou», cat. genoll, port. joelho;

esp. gente, cat. gent, it. gente, port. gente.

Enfin, on a, à l'évidence, une troisième formule de correspondances:

esp. -rm- = port. -rm- = cat. -rm- = fr -rm- = cors. rm = it. -rm-

qu'illustrent notamment les séries de mots:

esp. firme, port. firme, cat. ferm, fr. ferme, cors. fèrmu, it. fermo, roum. ferm; esp. fermento, port. fermento, cat. ferment, fr. ferment, cors. fermèntu, it. fermento, roum. ferment.

On voit donc que même les divergences lexicales dans des séries de mots peuvent mettre en oeuvre des formules de correspondances phoniques, et donc être finalement probantes pour qui entend établir la parenté des langues.

Cette divergence dans la désignation du frère s'explique aussi par un renouvellement du vocabulaire. En Ibérie, c'est un mot issu de lat. *germanus* qui a remplacé le mot issu de lat. *frater*, lorsque ce dernier s'est spécialisé dans la désignation du «frère en religion». Il s'agit donc du même problème que celui qui a entraîné en italien le remplacement de *frate* par *fratello* pour désigner le frère. Mais pour comprendre le renouvellement ibérique de *frater* par *germanus*, il faut savoir que ce dernier mot était, en latin, un adjectif qui signifiait «naturel, vrai, authentique» et s'employait notamment dans le syntagme:

frater germanus «frère germain»,

syntagme qui se réduisait ordinairement à l'adjectif substantivé:

germanus «frère germain».

C'est donc cet adjectif substantivé que le portugais, l'espagnol et le catalan ont secondairement utilisé pour signifier le «frère» tout court, lorsque celui-ci parut ne plus avoir de terme spécifique.

Il est intéressant, pour terminer ces considérations comparatives et génétiques, de noter d'une part que les parallélismes dans le vocabulaire et surtout les règles de correspondances phonétiques établissent incontestablement la parenté entre les toutes langues dites romanes, et d'autre part que leurs éventuelles divergences concernent souvent plusieurs

langues romanes et établissent ainsi une communauté plus grande entre ces langues semblablement divergentes.

**3. Les différentes langues romanes:** On peut proposer plusieurs types différents de classement des langues romanes.

**a. Classement politico-administratif:** On parle alors plutôt de langues latines. "La famille des langues latines comprend deux langues officielles des Nations Unies sur six: le français et l'espagnol, une autre langue parmi les dix les plus parlées au monde: le portugais, des langues nationales importantes, comme l'italien et le roumain, et quantité de langues ou dialectes d'importance diverse, tels que le catalan, le provençal, le corse, l'occitan, le galicien, le sarde, le romanche..." (Malherbe, Michel, 1983, *Les langues de l'humanité*, Paris, Seghers, 141).

**b. Classement littéraire:** Les romanistes qui s'intéressent à la linguistique comparée des langues romanes se limitent souvent aux six ou sept langues romanes qui sont des langues littéraires. Ce sont "le roumain, le rhétique, l'italien, l'ancien provençal, le français, l'espagnol et le portugais" (Meyer-Lübke, Wilhelm, 1890, *Grammaires des langues romanes*, trad. franç. par Eugène Rabié, Paris, Welter, 8).

**c. Classement géographico-dialectal:** Il serait plus intéressant d'avoir un classement en zones dialectales. On pourrait, en s'inspirant des différents classements possibles, mentionnés par Charles Camproux (cf. Camproux, Charles, 1974, *Les langues romanes*, *Que sais-je?* n° 1562, 77-78), proposer un classement en six grands domaines dialectaux.

•**le domaine ibéro-roman:** avec le portugais et l'espagnol. Ces deux langues sont très proches, si bien que, lorsqu'il n'y a pas de traitements phonétiques divergents, les comparatistes se contentent souvent de citer seulement l'espagnol.

"Les dialectes espagnols recouvrent le territoire politiquement espagnol excepté la zone catalane, la zone basque et la zone galicienne" (Camproux, 1974, 78). Ils se partagent "en trois groupes qui correspondent aux trois étapes de la «reconquista» sur les Arabes: au Nord, l'asturien, le léonais et l'aragonais; au Centre, le castillan (*castellano*), source de la langue littéraire <avec notamment le poème du Cid>; au Sud, l'andalou" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 50).

"Les dialectes portugais sont encore plus homogènes que les dialectes espagnols, bien que l'on distingue quatre zones dialectales. <...> Au groupe des parlers portugais se rattachent les parlers galiciens situés sur le territoire de l'ancien royaume de Galicie qui forme la partie nord-occidentale de l'Espagne" (Camproux, 1974, 79). Au Moyen Age s'est développée "une langue littéraire, le galicien-portugais —plus savamment appelé le gallaïco-portugais— dont le nom rappelle que jusqu'au XIV<sup>e</sup> siècle gallicien et portugais étaient restés une seule et

même langue" (Walter, 1994, *L'aventure des langues en occident*, 208). Cette langue littéraire fut la langue de prédilection de la poésie lyrique.

•**le domaine occitano-roman**: avec l'occitan, le gascon, le catalan. "Le groupe occitano-roman occupe une place centrale parmi les dialectes romans. On a pu parler à son propos de carrefour des langues romanes (Camproux, 1974, 81).

"Le catalan ancien différait assez peu de l'ancien occitan; il en différait, en tout cas, beaucoup moins que le gascon" (Camproux, 1974, 81). Il est parlé en France notamment dans le Roussillon, en république d'Andorre, dont il est la langue officielle, en Espagne, notamment en Catalogne (Barcelone) et dans la région de Valence, et dans une partie de la Sardaigne.

"Le gascon se subdivise en de nombreux sous-dialectes dont certains sont très proches de l'occitan languedocien et dont d'autres accumulent les traits originaux proprement aquitains. Parmi ceux-ci le béarnais a une très forte personnalité; il est resté langue écrite jusqu'au XVIII<sup>e</sup> siècle et son emploi officiel n'a entièrement disparu qu'à la Révolution française" (Camproux, 1974, 82).

L'occitan médiéval, appelé tantôt *lemosi* (limousin) tantôt *proensal* ou *proensales* (provençal), fut la langue des *troubadours*, qui "bien avant les *trouvères* de langue d'oïl et en marge des grands cycles épiques qui firent la gloire du nord de la France, <...> surent élaborer une poésie lyrique extrêmement savante, aussi bien par la forme que par le fond" (Bec, Pierre, 1978, *La langue occitane*, Que sais-je? n° 1059, (1ère éd.: 1963), 68). De fait "du XI<sup>e</sup> au XIII<sup>e</sup> siècle, l'occitan est vraiment la *langue type* de la poésie lyrique, comme plus tard le galaïco-portugais dans la péninsule Ibérique: les premières manifestations littéraires du français et du castillan étant plutôt épiques" (Bec, 1978, *La langue occitane*, 68).

•**le domaine gallo-roman**: avec le français et le franco-provençal, lequel forme en quelque sorte la "transition" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 49) entre le français et le provençal. La situation est en réalité un peu plus compliquée, car "les parlers franco-provençaux sont au contact des parlers italiens (Piémont), des parlers germaniques (Suisse alémanique), des parlers d'oïl et des parlers d'oc" (Camproux, 1974, 85).

•**le domaine rhéto-roman** ou roman alpin: il occupe "les vallées des Alpes comprises entre les Grisons <=en Suisse> et l'arc des Alpes Juliennes <=au nord de la Yougoslavie> jusqu'à l'Adriatique y compris encore à la fin du XIX<sup>e</sup> siècle, la ville de Trieste" (Camproux, 1974, 83). Il est constitué de trois groupes dialectaux isolés par la géographie même des Alpes, à savoir

–le groupe occidental des parlers Grisons (avec le romanche, qui "est depuis 1938 la quatrième langue officielle de la Confédération Helvétique" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 50), et l'engadinois),

–"le groupe central des dialectes ladins dans les Dolomites (Italie)" (Camproux, 1974, 84)

–"le groupe oriental du frioulan avec Udine <=ville du nord-est de l'Italie, à l'est de Venise> comme centre" (Camproux, 1974, 84).

•**le domaine italo-roman**: avec l'italien, le sarde, et, selon certains linguistes comme Tagliavini, le dalmate.

"A l'opposé des parlers espagnols et portugais, les parlers du groupe italien sont excessivement variés" (Camproux, 1974, 79); et à la différence des parlers français, "les dialectes italiens sont très vivaces" (Camproux, 1974, 79). On les divise généralement en trois grandes zones: les dialectes du Nord, avec notamment le piémontais, le ligurien et le vénitien, les dialectes du Centre, avec le toscan et le florentin, et les dialectes du Sud, avec notamment le napolitain et le sicilien.

Le sarde occupe évidemment le territoire de la Sardaigne. Mais les deux dialectes du nord de l'île, à savoir le parler de la région de Sassari et celui de la région de Gallura, sont des importations toscanes qui datent du XII<sup>e</sup> siècle. "Les deux principales variétés de dialectes proprement sardes sont le logoudorien dans le centre de l'île et le campidanien au sud. On considère le logoudorien comme le sarde par excellence: c'est le parler employé sous la dénomination de *sardo illustre* en littérature" (Camproux, 1974, 80-81).

Le dalmate fut parlé sur la côte yougoslave de l'Adriatique, depuis l'île de Veglia (aujourd'hui Krk) au nord-est de la Yougoslavie jusqu'au port de Raguse (aujourd'hui Dubrovnik, au sud de la Yougoslavie, en Croatie). "Le ragusain a dû disparaître vers le début du XVII<sup>e</sup> siècle" ((Meillet-Cohen, (éds), 1952, *Les langues du monde*, 50). Et "le dernier sujet vegliote parlant dalmate est mort en 1898. Le dalmate a été absorbé du côté roman par le vénitien, du côté slave par le croate" (Camproux, 1974, 84).

•**le domaine balkano-roman**: avec le roumain. Ce domaine va de la péninsule d'Istrie au nord de la Yougoslavie, c'est-à-dire sous Trieste et en face de Venise, jusqu'au bord de la Mer Noire, c'est-à-dire en Roumanie et en Bulgarie. Il comprend quatre principaux dialectes:

–l'istro-roumain en Yougoslavie,

–le méglenite, parlé dans la vallée de Meglena au nord-ouest de Salonique en Grèce

–le macédo-roumain ou aroumain, parlé surtout dans le nord de la Grèce (en Thessalie et en Macédoine), et aussi en Yougoslavie, Bulgarie et Albanie,

–le daco-roumain de la Roumanie proprement dite, "qui est essentiellement le roumain du nord du Danube" (Camproux, 1974, 84). "Le daco-roumain se subdivise <...> en valaque, base de la langue littéraire, au Sud; moldave à l'Est et dans quelques enclaves en Ukraine; et transylvanien à l'Ouest" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 51).

### Chap. III. Famille de familles de langues: langues indo-européennes.

La famille complexe des langues romanes est génétiquement apparentée à plusieurs autres familles de langues, avec lesquelles elle constitue la grande famille des langues dites indo-européennes.

Cette famille des langues indo-européennes comprend, à époque ancienne, la plupart des langues qui se parlaient en Europe et certaines langues d'Asie. Elle couvrait donc alors, comme son nom l'indique, un espace géographique qui allait de l'ouest de l'Europe jusqu'aux Indes. Par suite de l'histoire des peuples qui utilisaient ces langues, les langues indo-européennes en sont venues à être parlées, à époque moderne, "dans la presque totalité de l'Europe et de l'Amérique, dans une grande partie de l'Asie et de l'Océanie, dans une partie notable de l'Afrique" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 5).

Il est plus facile d'énumérer les quelques langues d'Europe qui n'appartiennent pas à la famille indo-européenne. Il s'agit premièrement des langues dites finno-ougriennes, avec au nord le finnois et l'estonien, et au centre de l'Europe, le hongrois; deuxièmement du basque, au sud-ouest de l'Europe; troisièmement du turc au sud-est de l'Europe; et quatrièmement, à l'extrême est de l'Europe, dans les montagnes du Caucase et les plaines avoisinantes, des langues caucasiennes comme au nord, en allant d'Est en Ouest, l'avar, le tchéchène et le tcherkesse, ou au Sud, le géorgien.

**1. Parenté génétique:** Fr. Bopp est le premier en 1847 qui ait établi avec précision la parenté et les caractéristiques propres des langues indo-européennes, en recourant à la méthode comparative. Cette parenté a naturellement été expliquée par le rattachement de toutes les langues ainsi comparées à une même langue-mère primitive, qui a été appelée l'indo-européen (en allemand *indogermanisch*). Cette langue-mère n'est pas attestée; c'est seulement une langue hypothétique qui a été reconstruite à partir de la comparaison entre les formes attestées dans les différentes langues indo-européennes historiques.

La parenté des langues indo-européennes s'établit à l'aide de règles de correspondances rigoureuses illustrées par un certain nombre de séries de mots comparables. On peut par exemple citer, comme pour les langues romanes, les noms de parenté suivant:

skr. *mātár-* «mère» (en fait: nom. *mātā*, acc. *mātár-am*), avest. *mātar-*, gr. μήτηρ (dor. μάτηρ), lat. *māter*, tokh. A *mācar*, arm. *mair*, v.irl. *māthir*, v.isl. *mó[r]*, vha. *muoter*, lett. *māte*, v.sl. *mati* (Gén. *matere*), lit. *móte* «femme mariée», alb. *motrë* «soeur (primitivement la soeur aînée, qui remplaçait la mère)» (d'après Krahe, 1966, 1, 36, et Ernout et Meillet, 1951, 694), à quoi on peut ajouter "v.nor. *mo[er]*, anglsax. *modor*, v.sax. *modar*, vha. *muoter*" (Krahe-Meid, 1969, 1, 52)

skr. pitā (acc. pitār-am), gr. πατήρ, lat. pater, got. fadar, (v.fries. fader, v.nor. fa[er, anglsax. faeder, v.sax. fadar, vha. fater), v.irl. athir, arm. hair (d'après Krahe, 1966, 1, 77 et Krahe, 1967, 50)

skr. bhrātar-, lat. frāter, v.irl. brāthir, got. brōthar, (v.fries. brōther, v.sax. brō[ar, anglsax. brō[or, ahd. bruoder, d'après Krahe, 1967, 31), arm. elbayr; "le slave et le baltique ont des dérivés: v.sl. bratrŭ (bratŭ), lit. broter-elis, etc." (Ernout-Meillet, 448); gr. ἀδελφός, mais φράτηρ «membre d'une phratrie (subdivision de la tribu)»

L'initiale des noms du «frère» illustre par exemple la règle de correspondances suivante:

skr. bh- = gr. ph- = lat. f- = germ. b- = sl. b- = lit. b- arm. b-

que l'on peut confirmer notamment par la fameuse série des verbes qui signifient «porter»:

skr. bhārāmi, av. barāmi, gr. φέρω, lat. ferō, got. bairan, (v.fries. bera, anglsax., v.sax. vha. beran, d'après Krahe, 1967, 53), arm. berem, v.irl. berim (-biur), v.sl. berō.

Cette règle de correspondances est ordinairement notée de façon raccourcie par le signe *\*bh*, qui représente donc toute la formule de correspondances, mais qui, en même temps, suggère que tous les sons attestés sont issus d'une sonore aspirée indo-européenne—plus précisément d'une occlusive bilabiale sonore aspirée—. Ceci voudrait par conséquent dire que la sonore aspirée de l'indo-européen s'est maintenue telle quelle en sanskrit, qu'elle s'est assourdie en grec, qu'elle est devenue une constrictive labiodentale sourde en latin, et qu'elle a perdu son aspiration dans les autres langues. On postule ainsi que tous les mots attestés remontent à une racine indo-européenne *\*bher-* «porter».

L'initiale des mots signifiant «père» et «mère» illustrerait les règles de correspondances respectivement de la bilabiale sourde indo-européenne *\*p* et de la nasale indo-européenne *\*m*. Plus intéressante est la consonne intérieure des noms du frère. Elle permet d'établir la règle de correspondances *\*t*, à savoir:

skr. -t- = gr. -t- = lat. -t- = germ. -θ- = sl. -t- = lit. -t- arm. -Ø- = irl -th-

que l'on pourrait également illustrer par:

skr. tráy-ah, gr. τρεῖς, lat. trēs, got. θrija (acc. neutre, cf. gr. τρία, lat. tria), v.sl. trije (rad. *\*trei-/tri-*).

Réduite à la correspondance entre l'indo-européen et les langues germaniques, cette règle de correspondances fait partie de ce que les manuels de germanistique appellent les "lois de Grimm". D'après en effet la première loi de Grimm, les occlusives sourdes indo-européennes correspondent en germanique à des fricatives sourdes, ou plutôt, en termes de changement phonétique, sont devenues en germanique des fricatives sourdes. C'est cette même loi de Grimm qui explique le passage de l'occlusive *\*p-* initial de l'indo-européen *\*pRter* à la fricative initiale de vha. *fater* (all. *Vater*).

Si maintenant on regarde ce qu'il en est dans les mots signifiant «père», on trouve exactement la même correspondance *\*t* dans toutes les langues indo-européennes sauf dans le

domaine germanique, où au lieu de la fricative -θ- attendue on observe une sonore, qui en gotique est notée par la lettre *d* de *fadar* —et c'est cette fricative sonore qui, par le renforcement articulatoire de ce qu'on appelle la seconde mutation consonantique, qui, elle, particularise l'allemand parmi les autres langues germaniques, va aboutir à la sourde [t] de vha. *fater*, tandis que parallèlement à la seconde mutation consonantique la constrictive sourde intervocalique de got. *brōθar* passe à l'occlusive sonore de vha. *bruoder* (all. *Bruder*). S'agit-il d'une anomalie, ou d'une sonorisation des fricatives intervocaliques qui se serait développée de façon sporadique seulement dans certains mots de certaines branches du germanique, alors qu'elle s'est généralisée de façon régulière dans d'autres branches comme le vieil anglais, où les deux mots ont un même [ʀ] intervocalique, comme en anglais moderne ['fa:ʀ] et ['brʋʀ]?

Verner, en recueillant tous les mots qui présentent ce traitement apparemment divergent, a pu montrer qu'ils correspondaient à une différence d'accentuation en sanskrit et en grec, où *pitá* et *πατήρ*, qui sont accentués sur la finale, s'opposent à *bhrátar-* et *φράτηρ*, accentués, eux, sur l'initiale. Cette nouvelle correspondance, qui avait l'air à première vue d'être une anomalie, reçut le nom de "loi de Verner", laquelle pourrait être formulée en ces termes:

"A l'initiale i.e. *p, t, k, k<sup>w</sup>* aboutissent toujours à des spirantes <=constrictives> sourdes, mais, placées entre deux éléments sonores, dont l'un est l'élément vocalique de la première syllabe du mot, elles deviennent des spirantes <= constrictives> sonores, sauf quand l'accent, hérité du ton indo-européen, portait, en prégermanique, sur cette syllabe" (Mossé et Jolivet, 1959, 53-54).

On objectera peut-être que le mot grec *μήτηρ* est accentué comme *φράτηρ* et a pourtant le même traitement que *πατήρ*. En fait il s'agit d'une évolution secondaire propre au grec; car en sanskrit ancien *mātár-* est accentué sur la finale. Il semble donc que le grec ait modifié l'accentuation primitive, qui était celle du sanskrit, en mettant finalement l'accent sur l'initiale, quand celle-ci était longue, *μήτηρ* au lieu d'un \**μητήρ* primitif, —ou sur la deuxième syllabe du mot, quand celle-ci était brève, comme dans *θυγάτηρ* au lieu de \**θυγατήρ* (cf. Mayrhofer, 1965, 26), parce que deux brèves sont l'équivalent d'une longue. Une telle évolution secondaire n'invalide donc pas la loi de Verner, qui s'applique bien sûr à l'accentuation grecque primitive, c'est-à-dire à l'époque où \**μητήρ* était accentué comme *πατήρ*.

**2. Les dialectes indo-européens:** Traditionnellement, les comparatistes indo-européanistes regroupent les langues indo-européennes en deux grandes sous-familles, à savoir les langues *kentum* et les langues *satem*. Cette dénomination vient de la façon de dire «cent» dans les différentes langues indo-européennes, à savoir:

i.e. \**kmtóm* : lat. *centum*, got. *hund*, v.irl. *cēt* (c'est-à-dire *kjēd*), gr. ἑ-κατόν (probablement *un-cent*), skr. *sātam* (*svatām*), avest. *śatām*, lit. *šimtas*, v.sl. *sūto*, tokhar. *kēnt(e)* " (d'après Szemerényi, 1970, 208).

Il y a donc un ensemble de langues indo-européennes qui présentent une sifflante initiale là où d'autres présentent une dorsale initiale, ce qu'on interprète ordinairement en disant qu'un certain nombre de dialectes indo-européens ont palatalisé la dorsale primitive, ce qui a finalement donné "s en russe comme en iranien, mais [š] en lituanien et une sifflante palatale en sanskrit, le *ch* de *ich* allemand" (Martinet, 64), toutes langues qui constituent donc les dialectes *satem*. C'est une évolution phonétique qui se reproduira plus tard dans les langues romanes, où toutefois la dorsale *k-* ne se palatalisera que devant les voyelles d'avant *e* et *i*. Les mots latins *centum* «cent» et *cinque* «cinq» (lat. cl. *quinque*) donnent en effet

port. cem [sem], esp. ciento [θjento], prov. cènt [sɛ̃], fr. cent [sɛ̃], cors. centu [tʰɛntu], it. cento [tʰɛnto], sar. kentü [kɛntu], alb. k'int [kjint], ~~rom.~~ sutã  
port. cinco [sinko], esp. cinco [θinko], prov. cinq [sin], fr. cinq [sɛ̃k], cors. cinque [tʰinkʷe], it. cinque [tʰinkʷe], sar. kimbe [kimbe], roum. cinci [tʰɪntʰɪ].

Apparemment la répartition géographique des langues *centum* et des langues *satem* est assez cohérente: les langues *centum* sont à l'ouest, et les langues *satem* à l'est. Parmi les langues *centum*, on a en effet les langues celtiques, avec le v.irl. *cēt* (c'est-à-dire *kjēd*), les langues germaniques avec le got. *hund*, où le \**k-* primitif a évolué en *h-*, mais où "ce passage de *k* à *h* n'avait pas commencé au VI<sup>e</sup> siècle avant notre ère, car le mot chanvre, qui s'est répandu en Europe vers cette époque et qui a les formes *kannabis*, *cannabis* en grec et en latin, a été emprunté en germanique assez tôt pour que le *k* puisse évoluer en *h* comme l'attestent les formes modernes de l'angl. *hemp*, all. *Hanf*" (Martinet, 1994, 65). Toujours parmi les langues *centum* se trouvent le lat. *centum*, et le gr. ἑ-κατόν. Par contre, l'iranien et le sanskrit, ainsi que le lituanien et le vieux slave, qui sont des dialectes *satem*, se trouvent, eux, à l'est.

Mais "cette répartition géographique assez cohérente s'est vue battue en brèche par la découverte et l'identification, au début du XX<sup>e</sup> siècle, dans le Turkestan chinois, d'une langue appelée au début, mais à tort tokharien. En dépit de cette localisation extrême-orientale, il s'agit d'une langue *centum* où «cent» est *kēnte*" (Martinet, 1994, 66). "Un peu plus tard, en 1917, le déchiffrement du hittite a révélé, en Asie mineure, la présence d'une autre langue *centum* où «cœur» se dit *kartis* avec le [k] du latin *cor*, du grec *kardia*, de l'irlandais *críde* et l'équivalent [h] régulier de l'angl. *heart* " (Martinet, 1994, 66). Pour André Martinet,

"La présence inattendue de langues *centum* à l'est du domaine pourrait s'expliquer, pour le hittite, en faisant valoir que cette langue et les autres langues anatoliennes se sont séparées du tronc commun avant que ne se manifestent les tendances à la palatalisation. Dans le cas du tokharien, pour lequel rien ne semble justifier une séparation aussi

ancienne, on peut supposer un déplacement d'ouest en est" (Martinet, 1994, *Des steppes aux océans*, 66).

On pourrait aussi invoquer ce que les comparatistes italiens ont appelé la "théorie des aires", selon laquelle une langue ou un dialecte nouveau se développe autour d'un centre culturel, social ou politique, constituant ainsi ce qu'on appelle une "*area maggiore*", et repousse à la périphérie sur les côtés la langue ou le dialecte ancien, lequel est ainsi cantonné dans des "*area laterale*" plus ou moins nécessairement disjointes.

### 3. Les familles de langues indo-européennes:

#### A. Les langues *satem*:

**a. La branche indo-iranienne** avec, à époque ancienne, le sanskrit et l'avestique: "C'est par cette branche de l'indo-européen que <l'on peut commencer, parce que> c'est la découverte du sanskrit, langue de l'Inde, par les occidentaux qui est à l'origine de la linguistique comparative" (Martinet, 1994, *Des steppes aux océans*, 68).

•**Langues indiennes**: "La forme la plus ancienne des langues indiennes est attestée dans le Rig-Veda, partie la plus archaïque des Vedas, textes sacrés du brahmanisme. La langue de ces textes est connue sous le nom de «sanskrit», en quelque sorte l'«élaborée». Elle s'est établie comme langue de culture de l'Inde tout entière, y compris des régions de langues dravidiennes <= autre grande famille de langues, mais non indo-européennes de l'Inde>. Les premiers textes sanskrits sont tardifs, mais ils nous offrent des formes linguistiques plus anciennes, transmises oralement, qu'on date volontiers d'avant -1000. Très tôt ont existé, concurremment avec la langue littéraire, ce qu'on désigne comme les prākritis, c'est-à-dire les parlers quotidiens plus évolués" (Martinet, 1994, *Des steppes aux océans*, 69).

"Aujourd'hui, il existe plusieurs langues indiennes comme le bengali, le goudjerati, le penjabi, mais la langue officielle de l'Etat indien, qui s'emploie concurremment avec l'anglais, est le hindi, forme littéraire de l'hindoustani, notée dans la graphie dite *devanagari*, qui sert également pour le sanskrit" (Martinet, 1994, 69).

•**Langues iraniennes**: "Les langues iraniennes les plus anciennement attestées sont le vieux-perse et la langue de l'Avesta. Le vieux-perse l'est par des inscriptions en graphie d'origine akkadienne, entre le VII<sup>e</sup> et le V<sup>e</sup> siècle avant notre ère. Il s'agit d'une forme occidentale de l'iranien, la Perse étant proprement la partie de l'Iran située près du Golfe Persique. C'était la langue des souverains dont les armées ont menacé la Grèce.

L'Avesta, bible du mazdéisme —[religion dualiste pour laquelle dans le théâtre du monde s'affrontent le principe du Mal et le principe du Bien; elle fut réformée par Zarathoustra (628-551 av. J.-C.), qui mit l'accent sur la transcendance divine et prêcha "une morale d'action fondée sur la certitude du triomphe de la justice" (LPLI, 1992, 1711)]— est un ensemble de textes religieux, de dates diverses et d'une interprétation linguistique délicate;

certaines parties sont rédigées dans une langue aussi archaïque que les plus anciens textes indiens" (Martinet, 1994, *Des steppes aux océans*, 69).

"Aujourd'hui, les langues iraniennes sont, essentiellement, outre le persan, langue officielle de l'Iran, le kurde, parlé en Iran, en Irak et en Turquie, le pas̄to qui est, en Afghanistan, en compétition avec le persan et déborde sur le Pakistan. Isolé dans le Caucase, se trouve l'ossète qui est peut-être ce qui demeure de la langue des anciens Scythes" (Martinet, 1994, *Des steppes aux océans*, 69).

Le trait linguistique qui caractérise le mieux l'indo-iranien ancien par rapport aux autres langues indo-européennes "est la confusion en *a*, long ou bref, de deux voyelles, distinguées ailleurs comme *e* et *o*, en grec, en italique et en celtique, comme *e* et *a*, dans d'autres langues" (Martinet, 1994, *Des steppes aux océans*, 68). On a pu le constater à l'occasion de la série de correspondances

skr. bhārāmi, av. barāmi, gr. φέρω, lat. ferō

où l'on a effectivement un *a* bref du sanskrit et de l'avestique qui correspond à un *e* bref du grec et du latin, et un *ā* long du sanskrit et de l'avestique, qui correspond à un *ō* long du grec et du latin.

**b. L'arménien:** "Parlé aujourd'hui encore en Arménie, là où l'Asie mineure confine au Caucase, l'arménien a essaimé du fait des persécutions dont ses usagers ont été l'objet. On trouve aujourd'hui des communautés arméniennes un peu partout, en URSS, en France, au Liban, aux Etats-Unis et ailleurs" (Martinet, 1994, 72).

"Les premiers manuscrits arméniens sont ceux d'une traduction de l'Évangile et datent du IX<sup>e</sup> siècle" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 36).

"On a cru, pendant quelque temps, que l'arménien était une langue iranienne du fait des nombreux emprunts faits par les Arméniens à leurs voisins du sud-est. La finale *-ian*, si caractéristique des noms arméniens, est en fait d'origine iranienne. En réalité, il s'agit d'une variété extrêmement typée de l'indo-européen, qui a fait une mutation consonantique analogue à celle qu'on retrouve en germanique <sans que ce parallélisme> implique <...> un apparemment particulier entre arménien et germanique" (Martinet, 1994, *Des steppes aux océans*, 72). On a vu en effet que les occlusives sourdes indo-européennes correspondaient à une constrictive aspirée ou à rien comme dans arm. *hair* «père» en face de gr. *patēr*, ou arm. *mair* «mère» en face de gr. *mētēr* (dor. *mātēr*), et que les occlusives sonores aspirées de l'indo-européen correspondaient à une occlusive sonore non aspirée comme dans arm. *berem* «je porte» en face de skr. *bhārāmi*, gr. *phérō*.

Des évolutions complexes rendent parfois peu apparents des rapprochements qui mettent néanmoins en œuvre des formules de correspondances rigoureuses. C'est ainsi que Meillet a établi que arm. *erku* «deux» était l'équivalent de i.-e. *\*dwō* (cf. lat. *duo* «deux», gr. homér. *δυω*), en établissant la correspondance systématique entre arm. *erk-* et i.e. *dw-*.

**c. L'albanais:** "Parmi les dialectes indo-européens, l'albanais est des plus tardivement connus: les premiers documents manuscrits sont une formule de baptême selon le rite romain (1642) et un petit vocabulaire recueilli par Arnold von Harff de Cologne (1496); les premiers textes imprimés sont le livre d'heures et le missel, traduits par Gjon Buzuk (1554-1555)" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 45).

On a souvent voulu rattacher l'albanais à l'ancien illyrien "sur la foi de la localisation géographique de l'Albanie. En effet, les Grecs situaient l'illyrie à peu près où se trouve l'Albanie d'aujourd'hui" (Martinet, 1994, 72). "Ce n'est pas certain, ni même qu'il descende d'une langue du même groupe" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 44).

**d. La branche balto-slave:** "Il s'agit là de langues *satem* caractérisées notamment par le fait que les <trois> séries d'occlusives établies à partir du sanskrit n'y sont représentées que par deux séries, une série /p/ et une série /b/. C'est ce qu'elles ont en commun avec l'iranien, mais elles s'en distinguent notamment en ce qu'elles ne confondent pas les timbres vocaliques [e] et [a], ni probablement, à date ancienne, [o] et [a]" (Martinet, 1994, 73). De fait, la série de mots signifiant «porter»:

skr. bhārāmi, av. barāmi, gr. φέρω, ..., v.sl. berō

nous a montré que la série des sonores aspirées *bh-* du sanskrit (*bhārāmi*) correspond simplement à une sonore *b-* en vieux slave (*berō*) comme du reste en avestique (*barāmi*), mais que le vieux slave distingue *e* et *o* (*berō*) à la différence de l'avestique qui ne connaît que le timbre *a* (*barāmi*).

•**Les langues Baltiques** sont attestées tardivement. Il s'agit du vieux prussien, du lituanien et du letton.

Le vieux prussien, "qui se parlait en Prusse orientale jusqu'au XVII<sup>e</sup> siècle" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 64), nous est principalement connu par deux catéchismes rédigés en 1545 "par des évangélistes germanophones, avec des graphies qu'il faut prononcer comme s'il s'agissait de l'allemand" (Martinet, 1994, 74).

Le lituanien et le lette, qui se parlent toujours, sont devenus, au XIX<sup>e</sup> siècle, des langues nationales et littéraires, puis redevenues, à la fin du XX<sup>e</sup> siècle, des langues nationales. Ils sont attestés depuis le XVI<sup>e</sup> siècle, "le lituanien par la traduction du catéchisme de Luther en 1547, le lette par la traduction d'un catéchisme catholique en 1585" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 65). Le lituanien est d'un archaïsme remarquable. Il a la particularité notamment de présenter "deux tons, sur les voyelles longues et les diphtongues accentuées" (Martinet, 1994, 74), l'un qui est descendant, noté par un accent aigu, l'autre, qui est montant, noté par l'accent dit circonflexe, qui est en fait un tilde, comme on l'a vu dans la diphtongue de lit. *šim̃tas* «cent».

•**Les langues slaves:** Le "vieux slave, ou slavon ecclésiastique ou «vieux bulgare»" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 65) est la langue de la

traduction de la Bible faite au IX<sup>e</sup> siècle par les deux frères, Cyrille et Méthode, qui, étant originaires de Salonique, ont évangélisé les Slaves. C'est la langue slave la plus anciennement connue; elle est encore la langue religieuse des Slaves orthodoxes.

Aujourd'hui "les langues slaves forment le groupe linguistique numériquement le plus important d'Europe. Elles se répartissent en trois groupes: méridional, occidental et oriental" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 65). Le groupe méridional, dont faisait partie le vieux slave, comprend le slovène, le serbo-croate et le bulgare. Le groupe occidental comprend surtout les parlers tchèques, les parlers slovaques, et le polonais. Le groupe "oriental comprend trois dialectes principaux dont la séparation n'est guère antérieure au XII<sup>e</sup> siècle" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 68), à savoir:

le grand-russe (*velikorússkiy*), qui est dénommé dans l'usage courant le russe, et dont "le dialecte moscovite a servi de base à la langue russe commune, littéraire et officielle" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 65),

le blanc-russe (*bielorússkiy*), qui est utilisé en Biélorussie,

et le petit russe (*rusniy ukráinskiy*), ruthène ou ukrainien, qui est la langue de la République d'Ukraine.

En ce qui concerne l'écriture, l'est du domaine slave, qui fut évangélisé à partir des Eglises d'Orient, dont la langue culturelle était le grec, utilise l'alphabet cyrillique. Il s'agit de tout le groupe oriental (grand-russe, blanc-russe et petit russe), et d'une partie seulement du groupe méridional (à savoir: le bulgare et le serbe). L'ouest du domaine slave, qui fut évangélisé à partir de l'Eglise d'Occident, dont la langue culturelle était le latin, utilise lui l'alphabet latin. Il s'agit de tout le groupe occidental (tchèque, slovaque et polonais), et du reste du groupe méridional (à savoir: le slovène et le croate). Faut-il rappeler que la séparation entre l'Eglise d'Occident et les Eglises d'Orient, qui fit suite à la séparation entre l'Empire romain d'Occident et l'Empire romain d'Orient, traversait l'actuelle Yougoslavie?

## **B. Les langues *centum*:**

**a. Le grec** est "la langue d'Europe dont les attestations écrites sont les plus anciennes" (Walter, 1994, *L'aventure des langues en Occident*, 46), et la langue qui fut le berceau de la civilisation et de la culture occidentales. "Dès les plus anciens témoignages, le grec apparaît déjà fractionné en dialectes" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 40), que l'on répartit, à époque ancienne, en groupe achéen, avec "l'arcadien, le cypriot et le pamphylien, attestés par des inscriptions" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 40), groupe éolien, groupe dorien, et surtout groupe ionien-attique. Ce dernier est "le plus important au point de vue littéraire: l'ionien, langue littéraire de la région qui a la première, dès le VII<sup>e</sup> siècle, développé une civilisation florissante; l'attique, langue archaïque et conservatrice d'Athènes, qui a produit aux V<sup>e</sup> et au VI<sup>e</sup> siècles une littérature qui

rayonne encore sur le monde civilisé" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 41).

"La langue du plus ancien monument littéraire, les poèmes homériques (IX<sup>e</sup>-VIII<sup>e</sup> s. pour les parties les plus anciennes), ne représente pas le premier état du grec: c'est un dialecte composite où, à un fond achéo-éolien, se sont superposés de nombreux traits ioniens. L'oeuvre a subi en outre une adaptation atticisante" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 41).

Au XX<sup>e</sup> siècle, la découverte en Crète, sur des tablettes, de documents de comptabilité écrits dans des langues dites linéaire A et linéaire B a fait remonter jusqu'au XV<sup>e</sup> siècle av. J.-C. les premières attestations du grec. Car "un Britannique, du nom de Ventris, qui, dans l'armée anglaise, avait servi dans le chiffre, a, en collaboration avec un helléniste, Chadwick" (Martinet, 1994, 82) établi que le linéaire B n'était qu'un état plus ancien du grec, désigné depuis sous le nom de mycénien.

Depuis le 1<sup>er</sup> siècle avant J.-C., il y avait en Grèce deux langues, l'une qualifiée de «purifiée» (*katharévousa*), qui sans être vraiment du grec ancien se rapprochait de lui et "différait du grec parlé, peu pour la phonétique, plus pour la morphologie et la syntaxe, plus encore pour le vocabulaire" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 42), l'autre qualifiée de «populaire» (*dimotiki*), qui était la langue grecque parlée. La constitution grecque de 1911 ne reconnaissait officiellement que la langue «purifiée». C'est seulement en 1976 qu'une loi reconnaît officiellement comme langue nationale la «langue néo-hellénique», c'est-à-dire, précise ce texte de loi, "la langue démotique, celle qui a été développée comme instrument d'expression panhellénique par le peuple grec et les écrivains reconnus de la Nation" (d'après Walter, 1994, *L'aventure des langues en Occident*, 1994, 51).

**b. Les langues italiques:** sont le latin, l'osque et l'ombrien, qui furent parlés en Italie. Les deux dernières langues italiques ont assez vite été absorbées par le latin, quand les peuples qui les parlaient ont été militairement battus par les Romains.

Le latin est la langue d'une bourgade du Latium, appelée Rome, dont l'importance a grandi au fur et à mesure des conquêtes que faisaient les habitants de cette petite ville. Il est ainsi devenu petit à petit la langue de l'Italie (272: prise de Tarente par les Romains), puis celle de l'ensemble du bassin méditerranéen (202: victoire romaine de Zama sur les Carthaginois) et d'une partie de l'Europe (118: fondation romaine de la ville gauloise de Narbonne; 52: siège d'Alésia par les troupes de César).

Le latin en tant que langue des Romains est attesté par des inscriptions et des textes littéraires entre 600 avant J.-C. et le 7<sup>ème</sup> siècle après J.-C.. Il cède ensuite la place aux différentes formes locales de latin parlé, qui sont à l'origine des langues romanes.

**c. Les langues celtiques:** que l'on a été tenté, probablement à tort, de réunir aux langues italiques en un grand groupe italo-celtique. On distingue deux groupes de dialectes celtiques, à savoir le celtique continental et le celtique insulaire.

Le celtique continental ou gaulois, qui occupait le centre de l'Europe, notamment la Gaule, l'Italie du Nord et l'Espagne, a été étouffé par le latin. Il en reste des noms propres principalement toponymiques, et une centaine de courtes inscriptions, difficilement interprétables.

Le celtique insulaire se subdivise en deux groupes: le gaélique d'abord, avec notamment l'irlandais, qui est, "selon la Constitution de la République d'Irlande, la «première» langue officielle" (Walter, 1994, *L'aventure des langues en Occident*, 80), mais qui, dans la vie quotidienne, est supplanté par l'anglais, la seconde langue officielle. C'est en gaélique d'Écosse que sont censées avoir été composées les œuvres du faux poète romantique Ossian, que James Macpherson prétendit avoir retrouvées dans les montagnes du nord de l'Écosse et qu'il traduisit en anglais en 1760. Le groupe brittonique en second lieu est formé du gallois, encore parlé au Pays de Galles, du cornique, qui a disparu au XVIII<sup>e</sup> s., et du breton, qui, chassé de Grande Bretagne (laquelle, à l'époque, s'appelait la Bretagne (*Britania*) tout court) par l'afflux des langues germaniques, a été introduit en Armorique aux V<sup>e</sup> et VI<sup>e</sup> s. après J.-C.

**d. Les langues germaniques:** sont, à l'intérieur des langues indo-européennes, caractérisées par la «mutation consonantique» de ce qu'on appelle les «lois de Grimm». Elles comprennent trois groupes de langues: le groupe oriental ou «ostique», le groupe septentrional ou «nordique», et le groupe occidental ou «westique».

- L'«ostique» est représenté par le gotique, la langue des Gots (Ostrogots et Wisigots), qui "est principalement connue par les fragments d'une traduction de la Bible que fit, pour l'édification de ses ouailles, l'évêque Wulfila ou Ulfila (né vers 311, mort vers 383)" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 57).

- Le «nordique», qui "est la langue germanique la plus anciennement attestée", grâce à des inscriptions du II<sup>e</sup> siècle "écrites en un alphabet dit «runique»" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 57), s'est scindé en quatre groupes entre le VIII<sup>e</sup> et le XI<sup>e</sup> siècle, à savoir l'islandais, le norvégien, le danois et le suédois. Le norvégien littéraire meurt au Moyen Âge, remplacé par le danois. Actuellement il existe en Norvège deux langues officielles: le *riksmål*, c'est-à-dire «langue du royaume», aussi appelé *bokmål* «langue du livre», et le *landsmål*, c'est-à-dire «langue de la campagne», aussi appelé *nynorsk* «nouveau norvégien». Le *riksmål* est très proche du danois, tandis que le *landsmål* "a été créé <...> en 1850 par Ivar Aasen à partir de la langue du Moyen Âge et des dialectes norvégiens en usage à cette époque" (Malherbe, 1983, *Les langages de l'humanité*, 156).

- Le «westique», qui est représenté principalement par l'allemand, comprend le haut-allemand (*hochdeutsch*), et le bas-allemand (*niederdeutsch*), réunis ordinairement sous l'étiquette générique de tudesque, et l'anglais.

"Au bas-allemand se rattachent le hollandais (*hollandsch*) et le flamand (*vlaamsch*), langues très voisines issues d'un mélange de dialectes bas-allemands apportés par les conquérants Francs et Saxons. Mais aujourd'hui c'est le hollandais qui est enseigné dans la

Belgique flamingante, sous le nom de néerlandais (*nederlandsch*)" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 60).

L'anglais est issu des dialectes westiques que parlaient les Angles, les Saxons et les Jutes, qui "venus du Nord de la Germanie <au cours des V<sup>e</sup> et VI<sup>e</sup> siècles ap. J.-C.>, envahirent successivement et conquièrent sur les Celtes bretons les parties orientales et méridionales de la Grande-Bretagne" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 61). On désigne sous le nom d'anglo-saxon, ou mieux de vieil anglais cet ensemble de dialectes, qui comprend "l'anglien, northumbrien au Nord et mercien au Centre, le saxon-occidental (west-saxon) au Sud et le kentien à l'extrémité Sud-Est" (Meillet-Cohen, (éds), 1952, *Les langues du monde*, 61).

Le haut-allemand, qui se distingue du reste du westique par ce qu'on appelle "la seconde mutation consonantique (*zweite* ou *hochdeutsche Lautverschiebung*) se divise, dans le temps, "en vieux-haut-allemand (vha.) du VII<sup>e</sup> au XI<sup>e</sup> siècle, moyen-haut-allemand (mha.) du XI<sup>e</sup> au XV<sup>e</sup> siècle, nouveau-haut-allemand (nha.) ou allemand moderne depuis le XV<sup>e</sup> siècle" (Jolivet-Mossé, 1959, *Manuel de l'allemand du moyen-âge (Des origines au XIV<sup>e</sup> siècle)*, 19). Le haut-allemand se subdivise géographiquement ou dialectalement en allemand moyen (*mitteldeutsch*) et allemand supérieur (*oberdeutsch*). L'allemand supérieur est l'allemand méridional, dont les principaux dialectes sont l'alémanique (en Alsace et en Suisse), le souabe et le bavarois. L'allemand moyen se trouvait au XII<sup>e</sup> siècle au centre du domaine allemand, c'est-à-dire au-dessus de l'allemand supérieur et en dessous de la ligne Cologne-Berlin qui séparait le haut-allemand du bas-allemand. Il comprenait les différentes formes du francique: le francique oriental, le francique rhénan et le francique moyen, auquel se rattache, en France, le francique lorrain. C'est dans le francique oriental du Thuringe et de la Saxe que Martin Luther, né en Thuringe, mais étudiant et professeur à l'Université de Wittenberg en Saxe, traduisit la Bible. Cette traduction de la Bible par Martin Luther, entreprise entre 1521 et 1534, est à l'origine de ce qu'on appelle l'allemand commun ou la «langue écrite» (*Schriftsprache*) allemande.

**e. Le "tokharien":** C'est en 1892 que "fut publiée la photographie d'un manuscrit rédigé dans une langue inconnue, extrait de la collection du consul russe à Kachgar (ou Kashgar, actuellement Kashi: 39.29° Nord / 76.02° Est), ville située à l'entrée du Turkestan chinois. Cette feuille ne fut identifiée que beaucoup plus tard, à la suite d'autres découvertes: elle contient plusieurs strophes d'un hymne au Buddha, dans la langue indo-européenne appelée «tokharien B»" (Pinault, Georges-Jean, 1989, "Introduction au tokharien", in: *Lalies* 7, 7). D'autres textes, très souvent tronqués, furent trouvés entre 1900 et 1915 dans la même langue et dans une autre qui fut appelée «tokharien B». Il s'agit moins de deux dialectes que de deux langues différentes, "entre lesquelles l'intercompréhension devait être impossible" (Pinault, 1989, 25). On les appela «tokharien» par une identification erronée du nom *twqry* (ou *toxri*) avec les Τόχαροι de Strabon, qui, en fait, "parlaient un dialecte moyen-iranien, le

bactrien" (Pinault, 1989, 24). Il semblerait que le «tokharien B» soit la langue de Koutcha, site où elle est seule attestée, et doit "en conséquence être appelé *koutchéen*. Quant au <tokharien> A, attesté uniquement par des textes provenant de la région de Qarachahr et de Tourfan, il a été identifié par H.W. Bailey avec la langue du royaume d'Agni, connu par les sources sanscrites, et il est appelé désormais *agnéen*" (Février, 1959, 350). Les textes de ces deux langues sont écrits dans des syllabaires "dérivés de l'écriture gupta" (Février, 1959, 349), laquelle est une des écritures indiennes septentrionales issues de l'écriture brāhmī.

La grosse majorité des textes en "tokharien" sont des textes religieux, d'inspiration bouddhique; un certain nombre sont de simples traductions de textes sanscrits, quelques-uns étant même des bilingues. Mais il y a aussi des textes relevant des sciences et des techniques, ainsi que des documents historiques, administratifs et juridiques.

**f. Le hittite:** Dernière langue indo-européenne découverte, le hittite est la langue utilisée dans les tablettes que Hugo Winckler découvrit près du petit village anatolien de Bogazköy, qui se trouve "à une centaine de kilomètres à l'est d'Ankara", la capitale de la Turquie, sur "un site qui devait se révéler être l'antique Hattoussa, capitale de l'empire hittite au cours du II<sup>e</sup> millénaire avant J.-C." (*Alpha Encyclopédie*, 3106). Ces tablettes étaient écrites en cunéiformes suméro-akkadiens. C'est en 1917 que l'orientaliste tchèque Bedrick Hrozny montra qu'il s'agissait d'une langue indo-européenne, et publia la première traduction de textes hittites. Avant le déchiffrement de ces textes, les hittites n'étaient connus que par la Bible et par des sources égyptiennes. On sait notamment qu'en 1272 av. J.-C. les égyptiens et les hittites livrèrent la fameuse bataille de Kadēš, célébrée en Egypte comme une victoire personnelle de Ramsès II. Mais cette victoire égyptienne fut aussi une victoire du roi hittite Chattušil III, qui, après celle-ci, conclut une alliance avec Ramsès II, auquel il donna sa fille comme épouse.

Une des premières choses qui mit Hrozny sur la piste fut d'interpréter le mot *vâdar*, qui s'opposait à l'idéogramme suméro-akkadien du pain, comme signifiant «eau». Il nota alors que "dans les inscriptions contenant le mot *vâdar* (nominatif-accusatif singulier) pour l'eau, je trouvais aussi des mots tels que *vedenaš*, gén., *vedeni*, dat.\_loc., *vedenaz*, abl., *vedenit*, instr. sing.", et que "en ces mots, d'après le sens de la phrase, j'étais forcé de reconnaître d'autres cas du mot *vedâr*" (Hrozny, 1931, "Le hittite: histoire et progrès du déchiffrement des textes", in: *Archiv orientální Journal of czechoslovak oriental institute*, 3, 278). Or l'indo-européen avait de vieux noms dont le signifiant présentait une curieuse alternance r/n, comme celui de l'«eau», qui se dit en gr. ὕδωρ, gén. ὕδατος (issu d'un ancien \*ὔδντος). Le sanskrit connaît les formes avec nasale: gén. *udnâh*, loc. *udân*, l'ombrien connaît les deux formes alternantes: nom.-acc. *utur*, abl. *une*. Le parallélisme d'une alternance aussi surprenante ne pouvait pas être dû au seul hasard; le hittite devait donc être apparenté aux langues indo-européennes. Hrozny procéda ensuite systématiquement à des comparaisons

entre les mots hittites et des mots indo-européens, et arriva ainsi à lire de plus en plus de tablettes, ce qui confirmait bien l'appartenance du hittite aux langues indo-européennes.

#### **Chap. IV: Autres grandes familles de langues.**

Il existe évidemment d'autres familles de langues comme la famille indo-européenne. Nous avons déjà signalé en Europe les langues finno-ougriennes (qui, avec les langues samoyèdes forment la grande famille des langues ouraliennes), les langues turques (qui, avec les langues mongoles et les langues toungouses, forment probablement la famille des langues altaïques) et les langues caucasiennes (à l'extrême est de l'Europe, dans les montagnes du Caucase et les plaines avoisinantes, avec au nord, en allant d'Est en Ouest, l'avar, le tchéchène et le tcherkesse, et au Sud, le géorgien).

**A. Familles bien étudiées:** Parmi ces autres familles de langues, il en est dont les données comparatives sont bien établies, et cela depuis longtemps. Meillet donne comme exemples "le sémitique, le finno-ougrien, l'indonésien, le bantou" (Meillet, "Introduction à la classification des langues", in: *Linguistique historique et linguistique générale*, II, 1952, 61).

**1. Les langues sémitiques:** La famille sémitique comprend 1) le sémitique oriental, 2) le sémitique occidental, lequel se subdivise à son tour en a) sémitique occidental septentrional, b) sémitique occidental méridional.

**a. le sémitique oriental** est représenté par l'akkadien, la plus ancienne de toutes les langues sémitiques, qui, en Mésopotamie, a supplanté le sumérien, mais a repris l'écriture cunéiforme que ce dernier avait inventée. L'akkadien a notamment été illustré par le Code du roi de Babylone (1793-1750 av. J.-C.) Hammourabi, lequel fut gravé sur une stèle conservée au Louvre.

**b. le sémitique (occidental) septentrional** regroupe deux branches dialectales, à savoir le cananéen et l'araméen.

• **Le cananéen** comprend:

- 1) le vieux cananéen des tablettes en cunéiformes de Tell el Amarna (15<sup>ème</sup> siècle av. J.-C.);
- 2) le phénicien principalement des inscriptions alphabétiques de Byblos (dont les plus nombreuses sont du 5<sup>ème</sup> siècle av. J.-C., et les plus anciennes des environs de l'an 1000 av. J.-C.), et celui de la ville de Carthage, appelé le punique, de quelques inscriptions et de 10 lignes de la comédie de Plaute, le *Poenulus*. C'est pour écrire cette langue que fut inventé un alphabet non cunéiforme exclusivement consonantique, appelé l'alphabet phénicien, qui est à l'origine de presque toutes les écritures alphabétiques modernes.
- 3) le moabite de l'inscription du roi Méša (9<sup>ème</sup> siècle av. J.-C.)
- 4) l'hébreu, la langue de la Bible et "de quelques rares et courtes inscriptions en caractères phéniciens (surtout: calendrier de Gezer en Judée, vers le IX<sup>e</sup> siècle; inscription du tunnel de Siloé à Jérusalem, vers la fin du VIII<sup>e</sup> siècle av. J.-C.)" (Meillet-Cohen, 1952, 113)

5) l'ougaritique des tablettes en cunéiformes alphabétiques de Ras-Shamra, que l'on peut dater du XIV<sup>ème</sup> siècle avant Jésus-Christ.

• **L'araméen** est subdivisé en araméen occidental et araméen oriental. L'araméen occidental comprend:

1) l'araméen ancien avec les inscriptions de Hama et de Zindjirli (8<sup>ème</sup> s. av. J.-C.), le nabatéen notamment de Pétra (du 1<sup>er</sup> siècle av. J.-C. au 1<sup>er</sup> siècle après J.-C.), le palmyrénien (du 1<sup>er</sup> siècle av. J.-C. au 3<sup>ème</sup> siècle après J.-C.), qui est la langue de Palmyre, capitale de caravaniers dans le désert de Syrie qui fut détruite par les Romains et dont la reine Zénobie participa enchaînée au triomphe de l'empereur Aurélien

2) l'araméen biblique du livre d'Esdras ou d'une partie du livre de Daniel

3) l'araméen de papyrus d'Égypte

4<sup>o</sup> l'araméen palestinien des targoum (traductions de la Bible) et du talmud (commentaires de la Bible) de Jérusalem

5) le samaritain (3<sup>ème</sup> et 4<sup>ème</sup> siècles ap. J.-C.)

L'araméen oriental comprend:

1) l'araméen juif du targoum et du talmud babylonien (4<sup>ème</sup> au 6<sup>ème</sup> siècles ap. J.-C.)

2) le mandéen des "chrétiens de Saint-Jean")

3) le syriaque de la ville d'Édesse (auj. Ourfa), métropole intellectuelle de l'Orient chrétien

4) le harranien de Harran, ville païenne au Sud d'Édesse.

**c. sémitique (occidental) méridional** comprend le nordarabique, le sudarabique et l'éthiopien:

1) Le nordarabique avec les inscriptions lihyanites (entre le 2<sup>ème</sup> ou le 1<sup>er</sup> siècle av. J.-C. et le 4<sup>ème</sup> ou le 5<sup>ème</sup> siècle ap. J.-C.) et les inscriptions thamoudiques, et surtout l'arabe du Coran, qui fut révélé au Prophète entre 612 et 632, et réuni en un «livre», après la mort de Mahomet, par son gendre 'Uthman (644-655), le troisième calife;

2) le sudarabique avec, à époque ancienne, les inscriptions minéennes ou sabéennes, et les dialectes modernes comme le mahri, le karawi et le sokotri;

3) l'éthiopien en Abyssinie avec notamment le guèze, le tigré et l'amharique.

**2. Les langues chamito-sémitiques:** On a proposé de considérer que les langues sémitiques ne représentent qu'une branche d'une famille de langues plus large, la famille des langues chamito-sémitiques. Les trois autres branches de cette vaste famille sont:

**a. l'égyptien:** avec l'égyptien ancien (de 4000 av. J.-C. au 7<sup>ème</sup> siècle ap. J.-C.), qui "ne survit que dans l'usage liturgique du copte, qui n'est plus parlé ni écrit" (Meillet-Cohen, 1952, 86), et le copte, langue avant tout des chrétiens d'Égypte ("le nom «copte» est une forme arabe, qui répond, moins la première syllabe, au grec *Aiguptos* " (Meillet-Cohen, 1952, 155)

**b. le libyco-berbère:** avec le libyque représenté par "de courtes inscriptions disséminées du Sinaï aux Canaries" et "encore mal interprétées" (Meillet-Cohen, 1952, 158),

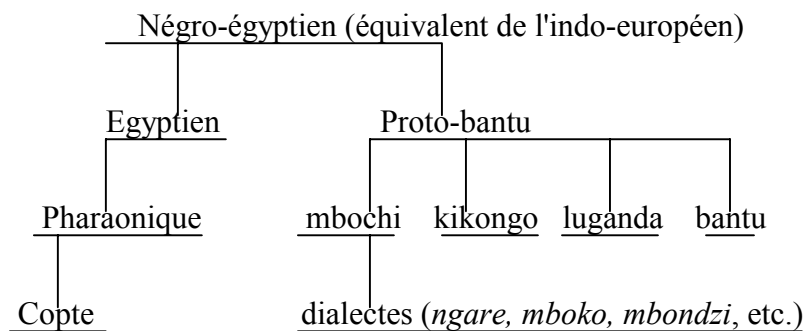
et le berbère, dont les dialectes sont "coupés en îles et îlots" (Meillet-Cohen, 1952, 86), avec notamment le berbère des Touaregs au Sahara, des Zenagra en Mauritanie, des Kabyles en Algérie, des Chleuh au Maroc.

**c. le couchitique:** avec notamment le méroïtique, l'afar, le somali. "Les langues couchétiques occupent presque toute la corne orientale de l'Afrique jusqu'à 4 degrés de latitude sud, en enrobant le domaine sémitique éthiopien à l'intérieur duquel elles forment encore des enclaves" (Meillet-Cohen, 1952, 165).

L'hypothèse de cette superfamille est récusée par l'africaniste sénégalais Cheikh Anta Diop, et surtout son disciple l'égyptologue Théophile Obenga, qui a cherché à établir que "les parallélismes entre le sémitique, l'égyptien et le berbère sont inexistants au niveau de la structure morphologiques des mots à une ou plusieurs consonnes radicales au niveau <...> des déclinaisons, <...> des pronoms personnels et relatifs, <...> de la conjugaison, <d>es particules de liaison (prépositions et conjonctions", bref, que "le «chamito-sémitique» ou «afro-asiatique» n'existe pas dans la matérialité des faits linguistiques. C'est une illusion, un mythe" (Obenga, 1993, *Origine commune de l'égyptien ancien, du copte et des langues négro-africaines modernes*, L'Harmattan, 95-96). Ils rapprochent par contre l'égyptien ancien des langues africaines.

"l'égyptien pharaonique, le copte, le bambara, le hausa, le wolof, le mbochi, le luganda, bref toutes les langues historiques parlées par les peuples noirs d'Afrique, depuis les temps obscurs de la proto-histoire, sont génétiquement apparentées entre elles: les faits morphologiques analysés ne disent pas autre chose" (Obenga, 1993, 159).

Et au terme d'une comparaison entre l'égyptien et le copte avec le mbosi, langue bantou parlée "dans la région de la Cuvette, en République du Congo" (Obenga, 1993, 24), qui est la langue maternelle de Théophile Obenga, ce dernier propose l'arbre généalogique suivant:



(Obenga, 1993, 249).

**3. "Les langues finno-ougriennes** sont: 1° Le lapon <...> parlé en Norvège et en Suède <...> en Finlande <...> et en U.R.S.S.

2° le finnois <avec> les dialectes suivants: a) *Suomi* ou finnois de Finlande, este, live b) carélien, vepse, vote

3° le mordve, 4° le tcherémisse, 5° les langues permiennes: le votiak et le zyriène, 6° les langues «ougriennes» de l'Ob: l'ostiak et le vogoul, 7° le hongrois ou magyar.

**4. Les langues ouraliennes:** On fait entrer la famille des langues finno-ougriennes dans une famille plus large encore, appelée les langues ouraliennes, qui sont donc parlées "depuis la côte septentrionale de la Norvège (Finmark) et les rives de la Lajta, affluent du Danube, jusqu'au delà du fleuve Iénissei, en Sibérie" (Meillet-Cohen (éds), *Les langues du monde*, 1952<sup>2</sup>, 279). La famille des langues ouraliennes réunit en effet le groupe des langues finno-ougriennes et le groupe des langues samoyèdes (cf. Meillet-Cohen (éds), *Les langues du monde*, 1952<sup>2</sup>, 280).

"Les langues samoyèdes, qui sont parlées à l'est de l'Oural, c'est-à-dire, à peu de choses près, entre les deux fleuves sibériens de l'Ob et de l'Iénissei, sont: 1° le yourak, <...> 2° le samoyède de l'Iénisséi, <...> 3° le samoyède de tavgui, <...> 4° le samoyède ostiak, <...> 5° le samoyède du Sud ou Sayan" (Meillet-Cohen (éds), *Les langues du monde*, 1952<sup>2</sup>, 286).

**5. Les langues bantoues:** "occupent toute l'Afrique centrale et l'Afrique méridionale sauf l'extrême coin sud-ouest " (848), soit presque la moitié de l'Afrique. Elles sont classées en 14 sections géographiques différentes, les plus connues étant le zoulou du groupe Ngouni de la Section du Sud-Est, le kinyaRwanda du groupe du Rwanda de la section du Nord-Est, et le groupe Swaheli de la section de la côte orientale.

**B. Regroupements encore plus larges?** Peut-on envisager des regroupements qui soient encore plus larges? Des tentatives ont été faites dans ce sens, mais elles sont considérées comme bien moins sûres que les regroupement proposés plus haut, et même sont assez souvent rejetées par les comparatistes spécialistes des familles indo-européennes, ouraliennes ou bantoues dont il a été question jusqu'à présent.

**1. Regroupement de l'indo-européen avec d'autres familles:**

**a. le nostratique:** "le linguiste danois Holger Pedersen proposa de rattacher l'indo-européen à d'autres familles telles que le sémitique, l'ouralien, l'altaïque et l'eskimo-aléoute au sein d'une famille qu'il appela nostratique" (Ruhlen, 1997, *L'origine des langues*, 83).

**b. On a aussi essayé de comparer le finno-ougrien avec l'indo-européen.** Et Alfredo Trombetti a écrit en 1905 (*L'Unità d'Origine del Linguaggio*, Bologne):

"Il est clair que la comparaison du finno-ougrien *me* «je, moi», *te* «tu, toi», avec l'indo-européen *me* et *te* vaut par elle-même exactement autant que toute comparaison que l'on peut faire entre les formes pronominales correspondantes dans les langues indo-européennes. La seule différence est que l'on reconnaît l'origine commune des langues indo-européennes, tandis que l'on rejette la relation entre l'indo-européen et le finno-ougrien" (d'après Ruhlen, 1997, *L'origine des langues*, 88).

**2. Le problème de la diversité des langues africaines:** "Lorsque Greenberg commença à faire des recherches sur les langues africaines, vers la fin de 1940, on dénombrait environ mille langues en Afrique, et leurs relations étaient en grande partie inconnues. Certaines des familles évidentes, telles que le bantou, le sémitique ou le tchadique, avaient été identifiées depuis longtemps, mais la majeure partie du continent restait *terra incognita*. Quand Greenberg publia sa classification, en 1963, il réduisait cette diversité déroutante à quatre familles en tout et pour tout: khoisan, nigéro-kordofanien (dont un sous-groupe était la famille bantoue), nilo-saharien et afro-asiatique (qui compte parmi ses branches le sémitique et le tchadique). Cette classification <...> constitue aujourd'hui la colonne vertébrale de toute linguistique africaine, et les africanistes, au bout de trente ans, considèrent que Greenberg n'a commis pratiquement aucune erreur dans sa classification, quoiqu'il eût à traiter des centaines de langues, dont beaucoup n'étaient que médiocrement décrites" (Ruhlen, 1997, *L'origine des langues*, 102).

**a. Opinion de Houis:** Voici par exemple l'opinion d'un africaniste français:

"il s'agit de réduire l'irritante diversité des langues. Pour Delafosse et Homburger, cette réduction est totale, puisqu'elle aboutit à poser une unité négro-africaine. Pour Greenberg, elle n'est que partielle puisqu'il propose quatre grandes familles: le nigéro-congo-kordofanien, le nilo-saharien, le khoisan, l'afro-asiatique" (Houis, 1971, *Anthropologie linguistique de l'Afrique noire*, 36).

"Greenberg ne se prononce pas sur l'unité négro-africaine. On remarquera seulement que, dans la première version de sa classification (1955), la réduction de l'ensemble était beaucoup moins poussée puisqu'il comprenait seize familles. Si l'on s'en tient strictement à la synthèse actuelle, la théorie de l'unité se trouve infirmée puisque l'ensemble se réduit à quatre grandes familles" (Houis, 1971, *Anthropologie linguistique de l'Afrique noire*, 37)

(cf. analyse de Maurice Houis du livre de J. H. Greenberg, in: *Bull. I.F.A.N.*, XXVI, B, 1-2, Dakar, 1964, 286-295).

**b. Les 4 familles d'après Greenberg** sont les suivantes, en allant du Sud vers le Nord:

le Khoisan, au sud de l'Afrique: en Namibie et en Afrique du Sud

le Nigéro-Kordofanien, bande médiane inférieure qui traverse l'Afrique d'Ouest en Est: depuis, à l'Ouest, la Guinée (où le Niger prend sa source) jusqu'au Mozambique au Sud-Est, en passant successivement, entre autres, par le Burkina Faso, le Nigeria et le Zaïre (où coulait le Congo, qui s'appelle maintenant le Zaïre)

le Nilo-Saharien, au centre droite de l'Afrique, au-dessus du Nigéro-Kordofanien: au Tchad, au Soudan et au Kenya

l'Afroasiatique, au-dessus du Nigéro-Kordofanien et du Nilo-Saharien: en Afrique du Nord, depuis la Mauritanie à l'Ouest jusqu'à l'Ethiopie et la Somalie dans la corne Est de l'Afrique.

Quand Merritt Ruhlen classe les grandes familles de langues du monde, il suit, en ce qui concerne les langues africaines, le classement de Greenberg, à cette différence près qu'il considère le Nigéro-Kordofanien et le Nilo-Saharien comme deux développements du Congo-Saharien. Greenberg, lui, répartissait entre le Nigéro-Kordofanien et le Nilo-Saharien les différentes langues qu'on pourrait désigner par l'appellation de Congo-Saharien.

Par ailleurs Merritt Ruhlen réunit en une même famille le Khoisan et les deux familles Nigéro-Kordofanienne et Nilo-Saharienne; mais il en sépare nettement l'Afroasiatique, qu'il situe parmi les langues non-africaines.

**3. Le problème de la diversité des langues d'Amérique:** "D'après certaines estimations, le nombre de <...> familles <de langues> indépendantes dans les Amériques approcherait de deux cents" (Ruhlen, 1997, *L'origine des langues*, 94).

**a. Le na-dene:** Mais Sapir a d'abord découvert la famille dite na-dene, qui réunirait la famille septentrionale des Montagnes Rocheuses et de l'Alaska dite famille athabasque et "trois autres langues parlées sur la côte méridionale de l'Alaska (Etat des Etats-Unis, au Nord-Ouest de l'Amérique septentrionale, qui est au-dessus et à l'Ouest du Canada), l'eyak, le tlingit (notamment dans l'Archipel de la Reine Charlotte) et le haida" (Ruhlen, 1997, *L'origine des langues*, 85; cf. Meillet-Cohen, 1952, 1026).

"D'autre part, en 1925, Sapir a annoncé qu'il avait pu établir la parenté de la grande famille Na-Dene avec la famille Sino-Thibétaine [Sapir, E., "The similarity of Chinese and Indian languages", in: *Science*, New York, n. s., t. 62, n° 1607, 16 octobre 1925, Supplément, p. XII]. Il a malheureusement été enlevé à la science avant d'avoir publié ses preuves" (Meillet-Cohen, 1952, 1026)

**b. Les 2 autres familles:** Puis Greenberg, dans *Language in the Americas*, (1987, Stanford University Press), a "démontré par de nombreux faits qu'il faut classer les langues aborigènes américaines en tout juste trois familles, et non dans les quelques deux cents sur lesquelles s'étaient mis d'accord les experts" (Ruhlen, 1997, *L'origine des langues*, 99), à savoir:

- "l'eskimo-aléoute, qui <...> est le plus oriental de la famille eurasiatique" (Ruhlen, 1997, *L'origine des langues*, 100)

- "la famille athabasque, qui constitue avec trois autres langues (l'eyak, le tlingit et le haida) la famille na-dene identifiée par Sapir en 1915" (Ruhlen, 1997, *L'origine des langues*, 100)

- l'amérinde, avec ses 11 branches: dont 4 "se trouvent presque exclusivement en Amérique du Nord" (Ruhlen, 1997, *L'origine des langues*, 100), à savoir: l'almosan-keresiou, le penutia, le hokan, l'amérinde central, et 7 "sont pour l'essentiel cantonnées à l'Amérique du

Sud" (Ruhlen, 1997, *L'origine des langues*, 100), à savoir: le chibcha-paez, l'andin, la macro-tucano, l'équatorial, le macro-caribe, le macro-pano, le macro-ge.

"Pour Greenberg et d'autres, les trois familles de langues autochtones américaines —l'amérindienne, le na-dene et l'esquimo-aléoute— dérivent de trois migrations séparées venues d'Asie" (Ruhlen, 1997, *L'origine des langues*, 111).

Il serait donc possible d'établir des parentés non plus entre des langues particulières, mais entre des proto-langues de langues particulières, c'est-à-dire finalement entre des familles déjà larges de langues.

**C. Les autres regroupements proposés par Joseph Greenberg:** Greenberg est ensuite allé beaucoup plus loin, en postulant deux autres grands groupes de langues.

**1. l'eurasiatique:** "Le premier grand groupe de familles <...> relie les familles indo-européenne, ouralienne, altaïque, coréenne-japonaise-aïnoue, tchouktchi-kamtchatkienne et esquimo-aléoute au sein d'une famille plus ancienne appelée eurasiatique par le linguiste américain Joseph Greenberg. Elle inclut aussi le guiliak (ou nivkh), parlé sur l'île de Sakhaline et sur la côte de Sibérie, lui faisant face" (Ruhlen, 1997, *L'origine des langues*, 81)

**2. le dene-caucasien:** "Le 2ème grand groupe de familles de langues <...> est connu sous le nom de dene-caucasien; il regroupe le basque, le caucasien, le burushaski, le sino-thibétain, le iénisséien et la famille américaine na-dene. Le na-dene se compose de la famille athabasque <...> et de trois langues parlées sur la côte méridionale de l'Alaska, l'eyak, le tlingit et le haida" (Ruhlen, 1997, *L'origine des langues*, 85).

"La distribution de dene-caucasien y est tout à fait frappante: ses membres apparaissent comme des îlots entourés de presque toutes parts par la vaste famille eurasiatique" (Ruhlen, 1997, *L'origine des langues*, 87)

**D. Les grandes familles de langues du monde:** Si on admet tous les regroupements envisagés par Greenberg, on peut, comme le propose Merritt Ruhlen, un disciple de Greenberg, dans sa carte des familles de langues du monde (cf. Ruhlen, 1997, 110), identifier 12 grandes familles différentes:

4 en Afrique:

- Khoisan
- Nigéro-kordofanien
- Nilo-saharien
- Afro-asiatique

2 en Europe, Asie et Amérique:

- Eurasiatique
- Dene-caucasien

1 en Amérique:

— Amérinde

5 dans le reste du monde:

— Australien

— Dravidien (en Inde)

— Austrique ou langues malayo-polynésiennes (Madagascar, Indonésie, sud du Vietnam, Nouvelle-Zélande)

— Kartvélien (en Géorgie)

— Indo-pacifique ou langues mélanésiennes (en Mélanésie et Micronésie).

**E. Monogénèse?** Peut-on aller plus loin et dire que toutes ces grandes familles de langues forment elles-mêmes une immense famille et sont par conséquent issues d'une seule et même proto-langue?

Un linguiste américain, Merritt Ruhlen, le pense. Dans un livre intitulé *On the Origin of Languages: Studies in Linguistic Taxonomy*, publié en 1994 par l'Université de Stanford, il a écrit en collaboration avec John D. Bengtson un article, intitulé "Global Etymologies", où, pour démontrer la monogénèse des langues existantes, sont proposées 27 étymologies qualifiées de mondiales. Ces étymologies ne permettent pas de reconstruire une proto-langue des différentes proto-langues déjà identifiées; mais elles montrent qu'il y a entre ces 27 mots ou proto-mots des ressemblances qui ne sauraient s'expliquer par un simple hasard. Et s'il n'est pas question de hasard, il s'agit alors d'héritage.

On peut illustrer ces ressemblances par les deux exemples de l'interrogatif «qui?», lequel, comme les linguistes le savent pertinemment, s'est, dans de nombreuses langues, développé aussi en indéfini «quelqu'un» et en relatif «qui». Citons comme exemple le pronom interrogatif latin *quis, quae, quid*, qui est quasiment identique à la variante *quis, qua, quid* du pronom indéfini *aliquis, aliqua, aliquid* «quelqu'un», et qui ressemble beaucoup au pronom relatif *qui, quae, quod* «qui».

#### «Qui?»

KHOISAN: *!kũ-(de)* «qui?», *!kung ka* «quand»

NILO-SAHARIEN: *na*, *didinga nani* «qui», *masaï ka* «quel»,

NIGÉRO-KORDOFANIEN: *nani*, proto-bantou *\*kí ~ \*ká*, swahili *ga-ni* «quoi, pourquoi»

AFRO-ASIATIQUE: *\*k(w) ~ \*q(w)*, proto-sémitique *\*kV* «comment», arabe *ka*, akkadien *kima ~ ki* «comment», proto-couchitique *\*kw* «qui», somali *kú-ma* «qui (masc.)», proto-tchadique *\*k'(w)* «qui», haoussa *k'ā*

CAUCASIEN: proto-caucasien *\*kʷi* «qui, quel», avar *kʰi-n* «comment»

NA-DENE: haida *gyis-to* ~ *kiš-to* «qui», *gōsu* ~ *guu(s)* «quoi», tlingit *kusu* ~ *gùsú* ~ *gū* «où»

DRAVIDIEN: *yāv*

EURASIATIQUE: proto-indo-européen *\*k<sup>w</sup>o* ~ *\*k<sup>w</sup>i* «qui», proto-ouralien *\*ke* ~ *\*ki* «qui», *\*ku* ~ *\*ko* «qui, quel», proto-türk *\*k'Em* «qui», *\*ka* ~ *\*qa* (base interrogative), ancien turc *kām* «qui», *qa-ñu* «quel», *qa-na* «où», coréen *ka* (particule interrogative), japonais *ka* (particule interrogative), proto-eskimo-aléoute *\*ken* «qui», *\*qa-* (pronom interrogatif)

AUSTRO-ASIATIQUE: munda *o-ko-e* «qui», *o-ka* «quoi»

?AUSTRALIEN: maung *gunuga* ~ *gigi* «quoi», tiwi *kuwa* «qui», *kamu* «quoi»

AMERINDE: seneca *kwanu* «qui», tunica *kaku* «qui», natchez *kanne* «quelqu'un», proto-aztèque *\*kaan* «où»

#### «eau»

KHOISAN: !o!kung *kāũ* «pleuvoir», !kung *k'ā* «boire»,

NILO-SAHARIEN: fur *kí* «pluie», nyimang *kwe* «eau»

NIGÉRO-KORDOFANIEN: *engi* «qui»

AFRO-ASIATIQUE: proto-afro-asiatique (Illitch-Svitytch) *\*šq(w)* «eau», proto-couchitique *\*-k'w-* «être humide»

CAUCASIEN: proto-caucasien *\*-VqV* «sucrer», proto-lesghien *\*Foχ<sup>w</sup>a-* «pluie», tchéchéne *-aq-* «sucrer»

DRAVIDIEN: *nīru* «eau»

EURASIATIQUE: proto-indo-européen *\*ak<sup>w</sup>ā-* «eau», hittite *eku-* «boire», latin *akwa* «eau», gotique *ahwa* «rivière», proto-ouralien (Rédei) *\*yoka* «rivière», japonais *aka* «eau de cale»

AUSTRIQUE: *namaw* «eau»

INDO-PACIFIQUE: awyu *okho* «eau, rivière», yareba *ogo*, yonggom *oq*

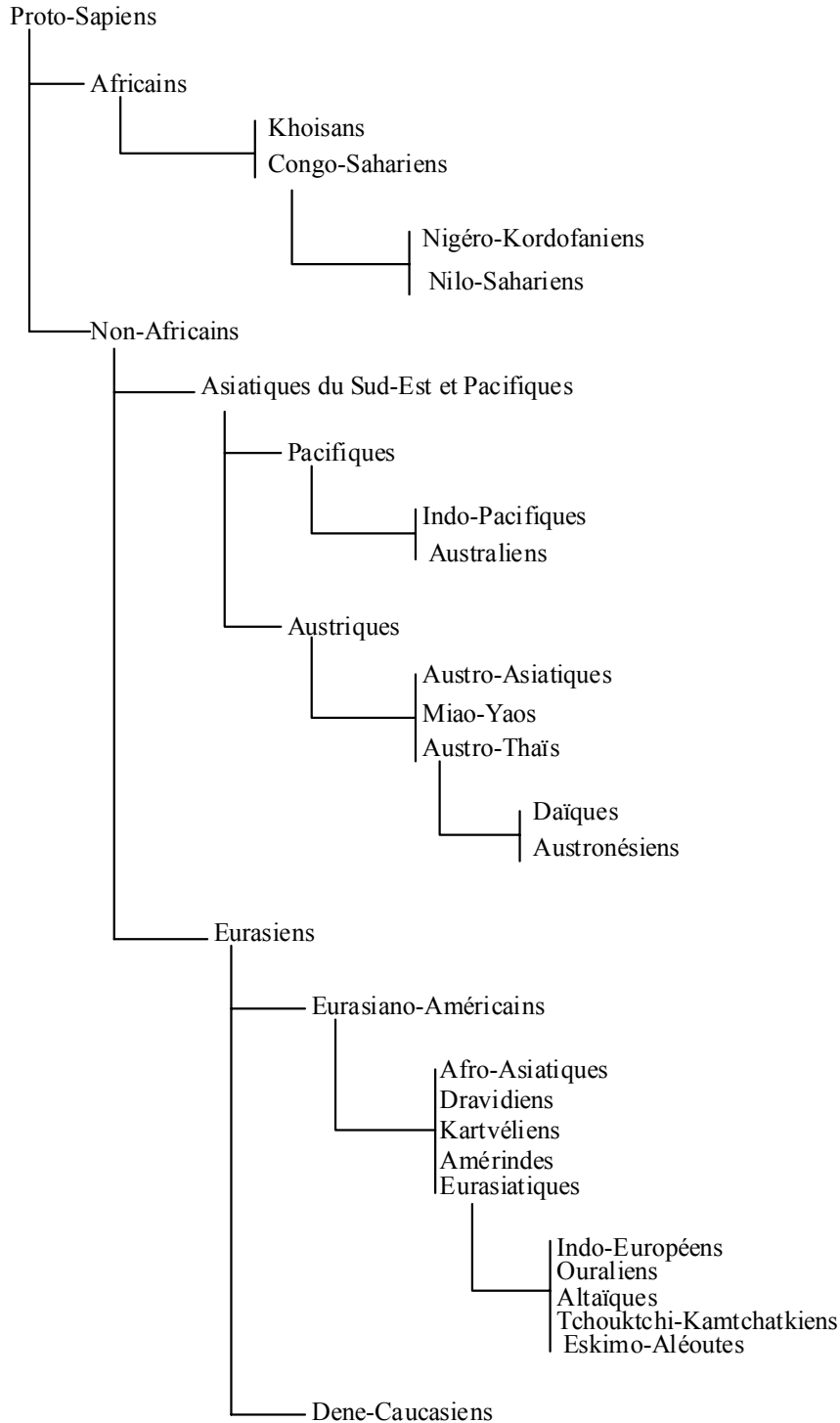
AUSTRALIEN: proto-australien *\*gugu* «eau»

AMERINDE: proto-algonquin *akwā* «hors de l'eau», proto-chinantèque *\*g<sup>w</sup>a* «courant, rivière», quechua *yaku* «lac»

(d'après Merritt Ruhlen, 1997, *L'origine des langues, Sur les traces de la langue mère*, trad. par Pierre Bancel, (original: 1994), Paris, Belin.

Pensant avoir justifié l'hypothèse que "toutes les langues du monde sont apparentées" (Ruhlen, 162), Merritt Ruhlen s'efforce ensuite de "situer ces conclusions linguistiques <...> dans le contexte plus large de l'évolution humaine" (Ruhlen, 163). Il constate que les regroupements de famille de langues proposés par les linguistes recourent plus ou moins les hypothèses génétiques que font actuellement les biologistes en ce qui concerne le peuplement de l'humanité. Il discute la comparaison qu'avait faite Luca Cavalli-Sforza "entre les groupes de populations (identifiés sur la base des fréquences des gènes contenus dans les noyaux cellulaires) et les familles de langues découvertes par des linguistes

tels que que Greenberg" (Ruhlen, 1965). Et finalement, "en reprenant des étiquettes linguistiques pour les différentes populations, il <reformule> la structure génétique des populations humaines élaborée par Cavalli-Sforza et ses collaborateurs dans le diagramme" (Ruhlen, 1970) suivant:



La structure généalogique de la population humaine  
(d'après Merritt Ruhlen, 1997, *L'origine des langues*, 211)

## BIBLIOGRAPHIE

- Bec, Pierre, 1978<sup>4</sup>, *La langue occitane*, (1ère éd.: 1963), Paris, PUF, Que sais-je? n° 1059, 128p.
- Bouvier, Jean-Claude, 1979, "L'occitan en Provence, Le dialecte provençal, ses limites et ses variétés", in: *Revue de linguistique romane*, 43, 46-62.
- Camproux, Charles, 1974, *Les langues romanes*, Paris, PUF, Que sais-je? n° 1562, 128p.
- Fishman, Joshua A., 1968, "La situation linguistique aux Etats-Unis", in: Martinet, André, (éd.), 1968, *Le langage*, Encyclopédie de la Pléiade, Paris, Gallimard, 1203-1221
- Fourquet, Jean, 1968, "La situation linguistique en Allemagne", in: Martinet, André, (éd.), 1968, *Le langage*, Encyclopédie de la Pléiade, Paris, Gallimard, 1162-1172.
- Guiraud, Pierre, 1968, *Patois et dialectes français*, Que sais-je? n° 1285, 128p.
- Kersaudy, Georges, 2001, *Langues sans frontières. A la découverte des langues de l'Europe*, Editions Autrement, 383p.
- Malherbe, Michel, 1983, *Les langages de l'humanité, Une encyclopédie des 3000 langues parlées dans le monde*, Paris, Seghers, 444p.
- Martinet, André, (éd.), 1968, *Le langage*, Encyclopédie de la Pléiade, Paris, Gallimard, 1525p.
- Martinet, André, 1994, *Des steppes aux océans, L'indo-européen et les «Indo-Européens»*, Paris, Payot, 274p.
- Meillet, Antoine, et Cohen, Marcel, 1952<sup>2</sup>, *Les langues du monde*, Paris, CNRS/Champion, 21 cartes + 1296p.
- Pottier, Bernard, 1968, "La situation linguistique en France", in: Martinet, André, (éd.), 1968, *Le langage*, Encyclopédie de la Pléiade, Paris, Gallimard, 1144-1161.
- Ruhlen, Merritt, 1997, *L'origine des langues*, trad. de l'anglais par Pierre Bancel, (original: 1994), Paris, Belin, 288p.
- Walter, Henriette, 1994, *L'aventure des langues en occident (Leur origine, leur histoire, leur géographie)*, Paris, Laffont, 498p.