



Mai 2006

1. Introduction

Continuel sujet de perplexité pour les linguistes : comment, d'une même faculté de langage propre aux humains, peut jaillir l'immense diversité des langues ? 6000 sont en effet parlées sur la planète. Cette question resterait à jamais insoluble sans un recours aux autres sciences. Si un coin du voile commence à être levé, c'est que l'archéologie, la paléanthropologie, la génétique des populations, la psychoneurologie, l'éthologie et d'autres disciplines viennent, avec la linguistique, instruire le dossier. Pas à pas, elles analysent le présent et remontent le temps.

2. Où commence la diversité ? Le langage et les langues

La faculté de langage est inscrite au génome humain, sans qu'il y ait à proprement parler de « gènes du langage », mais simplement une implication de certains gènes dans des processus qui culminent dans le langage. Elle est considérée aujourd'hui par les neuropsychologues comme reposant sur la spécialisation d'aires cérébrales qui contribuent par ailleurs à d'autres aspects de la cognition. Grâce notamment aux techniques de l'imagerie cérébrale, on peut entrevoir qu'il existe une grande hétérogénéité dans l'exploitation de la faculté de langage : complexité et flexibilité en sont les caractéristiques principales. Et ce qui est observable sur les populations actuelles peut, dans une certaine mesure, être extrapolé sur des *Homo sapiens* un peu plus anciens, puisque leurs capacités langagières semblent avoir été similaires (Mandel, 2003, p. 48 ; Hublin, 2003, p. 26, 2005, p. 116 ; Nespoulous, 2005, pp. 440, 447-448). Selon l'hypothèse la plus vraisemblable des paléanthropologues, cette aptitude de langage moderne se serait mise en place par paliers successifs, soit dès l'émergence de notre espèce, il y a environ 150 000 ans, soit lors de l'explosion culturelle il y a 50 000 ans (Coupé, 2005, p. 169). Il se trouve que le canal visuo-gestuel et le canal vocal-auditif sont également susceptibles de construire du sens. Mais, par avantage adaptatif, un langage articulé à caractère vocal pourrait avoir émergé à travers la longue durée (voir Annexes 1, Encadré 2 : « Quatre domaines de divergence »), à partir d'un protolangage rudimentaire de *Homo ergaster* et de *Homo erectus*. Ce protolangage se présenterait comme un système de communication peu performant, en étroite dépendance avec le « ici et maintenant » (Hublin, 2003, p. 30 ; Hombert, 2005, pp 462-463). Dès lors, la nette distinction opérée par un noyau de linguistes tels que André Martinet (1968, p. XI) et Claude Hagège (2003, p. 154) entre aptitude au langage et l'une de ses cristallisations : les langues, apparaît comme pleinement justifiée. Les généticiens des populations et les paléodémographes, pour leur part, nous apprennent que les premiers peuplements d'*Homo sapiens* étaient très réduits, les espaces géographiques probablement restreints, et les interactions humaines faibles. Ce qui présuppose une diversité linguistique elle aussi très ténue (Coupé, 2005, p. 183). Ainsi, notre dispositif cérébral tout comme les conditions d'évolution des *Sapiens* semblent favoriser l'éclosion de la diversité. Avec cette toile de fond éclairante, le linguiste peut, à son tour, apporter sa pièce à l'édifice.

3. Langues, dialectes et idiolectes

C'est un fait notoire que les 6000 langues planétaires se subdivisent en unités plus petites, les dialectes et sous-dialectes, dont les locuteurs ne se comprennent pas toujours entre eux. La distinction entre langue et dialecte n'étant pas toujours claire, ne serait-ce qu'en raison de l'intervention des États. Le mandarin est, en réalité, un dialecte officialisé. Il en va de même du français, issu du « francien » naguère parlé dans le Bassin parisien. Dans bien des cas, on baptise « langue », à l'inverse, une abstraction représentant un continuum de parlers et de dialectes sur un espace donné. Il en va ainsi du tcherkesse, au Caucase, ou du quetchua,

l'ancienne langue des Incas. Les dialectes se fragmentent à leur tour en unités encore plus petites qui aboutissent aux idiolectes, systèmes linguistiques propres aux individus. Les abstractions « langue » et « dialecte » ne sauraient conduire à une réification de ces concepts pour tous ceux qui acceptent de les voir de manière dynamique comme un perpétuel ajustement d'idiolectes (premiers degrés des contacts linguistiques), indéfiniment combiné avec une constante application du code, ou une dialectique typique des comportements humains. Comme l'énonce Salikoko S. Mufwene (2001, p. 2) :

To be sure, a communal language is an abstraction inferred by the observer. It is an extrapolation from I [internalized] –language whose speakers communicate successfully with each other most of the time.

Ce n'est que par abus de langage métaphorique (« guerre des langues », « langues dominantes », « langues dominées » etc.) qu'on peut être amené à les considérer comme des entités évoluant indépendamment des sujets parlants.

4. Vie, mort et renaissance des langues

Chacun sait ou croit savoir, par ailleurs, que les langues « naissent », « meurent », et occasionnellement « renaissent ». Mais c'est encore une métaphore. L'une des multiples que les linguistes utilisent par emprunt à la biologie, en l'occurrence au vitalisme du XIXe siècle, par analogie avec les espèces vivantes, ou en puisant dans le fonds religieux, tout en gardant à l'esprit les limites de cette transposition. Elle ne doit pas faire perdre de vue que le second cas (la mort) est plus facile à détecter que le premier. La métaphore biologique a d'ailleurs ses failles : une même langue peut mourir dans un lieu et survivre dans un autre. Le français est mort dans le Maine mais pas au Québec ; le yidiche est mort en Alsace, mais vit en Israël, en Argentine ou à New York. Une nouvelle langue peut apparaître par scission au sein d'une langue donnée, et dans ce cas échapper à la conscience de plusieurs communautés de locuteurs, dont les usages divergent peu à peu jusqu'à franchir une limite, au-delà de laquelle ils cessent de se comprendre. Au début de ce siècle, Antoine Meillet (1965, réédition), aimait à rappeler qu'à la fin de l'Empire romain, une longue chaîne d'hommes, dont chacun avait eu le sentiment de parler latin, était en fait en train de donner naissance à toute la gamme des langues romanes.

Dans un passé récent et aujourd'hui même, nous assistons dans plusieurs régions du monde à d'autres formes de « naissance » : les créoles (au sens large, typologique du terme) et les pidgins. Les créoles (au sens étroit, socio-historique du terme, cf. Hazaël-Massieux, 2005, pp. 3-17), nés au cours des colonisations européennes des XVIIe-XVIIIe siècles, sont le parler de descendants d'esclaves noirs mis en contact avec la langue de leurs maîtres, anglais ou français par exemple. Par extension, on peut dénommer créoles et créolisation des langues ainsi qu'un processus qui impliquent une dynamique de contacts très intenses, mais ces concepts ne font pas l'unanimité. Dans cette perspective, le français d'Abidjan, le joual, ou le cap-verdien sont des créoles, voire pour certains qui poussent le trait, l'anglais ou le chinois. Une telle position est notamment combattue par Robert Chaudenson (2000) et Mufwene (1986, 2001). Les pidgins, attestés historiquement avant les créoles, sont le parler plus rudimentaire de communautés de langue différente mais forcées de commercer entre elles. Leurs systèmes sont plus simples, leur fonction de communication restreinte, et en outre, ces parlers n'ont pas le statut de langues maternelles. Toutefois, dans certains cas, ils évoluent vers des pidgins étendus (« expanded pidgins »), véritables *lingua franca*, plus complexes, et parfois même natives.

Ils sont dès lors analogues aux créoles (cf. Mufwene, 2001, p. 7). Quoi qu'il en soit, il s'agit, dans les deux cas, de la manifestation extrême d'une situation, où le contact prolongé entre deux ou plusieurs langues conduit à une réorganisation complète de divers systèmes linguistiques. Comme l'énonce encore Mufwene (2001, p. 192) :

Creole vernaculars are not outcomes of abnormal, unusual, or unnatural developments in language evolution. Rather, they make more evident restructuring processes that must have taken place in the evolution of other languages.

En outre, vifs sont les débats autour de l'ontogénèse et de la phylogénèse d'un côté (cf. Bickerton, 1981), et des substrats de l'autre, suscités par cette situation d'observatoire en quelque sorte. Il se produit, comme le décrit Hagège (1985, pp. 29-41), sur fond, encore mal cerné, des facultés cognitives propres à l'espèce humaine. A un autre extrême, plus rare, il

arrive qu'une langue « morte » renaisse du fait d'une volonté politique et idéologique. Le cas le plus spectaculaire est l'hébreu moderne. La situation quasi-expérimentale de cet idiome ayant cessé d'être vernaculaire depuis au moins 1 500 ans et réanimé par un puissant volontarisme oblige l'observateur à un recentrage vers l'homme architecte de sa langue. Elle s'inscrit dans un contexte historique qui l'autorise (nationalisme et création d'un État) et un contexte culturel (courant « millénariste » et courant « esthétique » hérité de l'époque des Lumières) (cf. Szulmajster-Celnikier, 1997, pp. 141-142, 145).

4.1. Le nombre de langues : équilibre et rupture

Qu'en était-il dans le lointain passé ? Robert M.-W. Dixon (1997) affirme, en s'inspirant de la biologie évolutionniste, que l'histoire de la diversité des langues, à un niveau macroscopique, a été ponctuée de longues périodes d'équilibre, au cours desquelles le nombre de langues était relativement constant ou en augmentation notable, entrecoupées de phases de rupture qui voyaient mourir de nombreuses langues et en naître d'autres. La première phase d'équilibre aurait duré plus de 40 000 ans, de -50 000 à -10 000 avant le présent, à la fin du Paléolithique supérieur (cité par Peyraube, 2005, pp. 452-53). Daniel Nettle (2002) évalue globalement, sur la base de considérations démographiques, le nombre de langues d'alors entre 1 667 et 9 000. La première période de rupture remonte au début du Néolithique, lorsqu'apparaissent, en divers points du globe de manière parallèle, les premières sociétés d'agriculture et d'élevage, soit entre -8 500 et -3 000 ans. Des langues en remplacent d'autres, mais leur nombre demeure à peu près stable. Une deuxième période d'équilibre s'est alors instaurée, qui a perduré jusqu'à un passé récent, celui des grandes conquêtes maritimes européennes et chinoises du XVe siècle. Une poussée démographique a eu, à cette époque, pour corollaire un accroissement considérable de la diversité linguistique. Il est important de souligner que pour certains paléoanthropologues, éthologues et sémiologues, les premières traversées maritimes des *Sapiens* pourraient constituer un indice comportemental de l'émergence du langage. Au-delà de la préhistoire, les conquêtes maritimes pourraient être interprétées comme des événements susceptibles de changer la direction d'une trajectoire évolutive (en termes mathématiques, comme les « attracteurs » de la théorie du chaos). Toujours est-il que cette phase d'équilibre a probablement cessé aux XVe et XVIe siècles, période de la seconde rupture, toujours en cours. Par conséquent, le nombre de langues n'aurait cessé de croître jusqu'au XVIe siècle, pour commencer à décliner à partir de cette date (Peyraube, 2005, p. 455).

4.2. Les causes d'extinction

Les causes de la mort des langues sont aussi diverses que celles qui président à leur naissance. Certaines disparues sont demeurées célèbres en raison de l'impact de leur civilisation (cas du sumérien), de l'intérêt de leur écriture (l'étrusque) ou simplement de leur découverte récente (le hittite, voir Annexes 2, « Photo de divers types d'écriture »). Bien plus nombreuses, sont celles qui se sont évanouies sans laisser de trace. La cause d'extinction la plus courante est le contact prolongé entre deux langues. On a vu qu'il peut aussi être à la source de nombreux idiomes, et qu'il constitue l'un des moteurs du changement linguistique. Mais il peut apparaître encore comme un facteur d'acculturation, insidieuse ou agressive. Les langues disparaissent ainsi à la suite de processus de transformation, de substitution et d'extinction (Hagège, 2000, pp. 93-94). Les raisons profondes de cette disparition sont le plus souvent d'origine extralinguistique : politiques, économiques, culturelles, sociales ou liées à des catastrophes naturelles, ou provoquées par l'homme (Hagège, 2000, chap. VII).

4.3. Un crescendo dans l'extinction ?

Le phénomène semble s'être accéléré depuis le XVIe siècle, d'abord en raison de l'expansion et de la domination européennes, puis, plus récemment, sous la pression de la mondialisation des échanges et de l'urbanisation de populations rurales et dispersées. Selon les estimations, quelques centaines, voire plus d'un millier de langues risquent de mourir à court ou moyen terme. Pour Hagège (2000), une langue disparaît tous les quinze jours. À ce rythme, dans un siècle, la moitié de notre patrimoine pourrait s'évanouir. Exit le cupeño, le matipu, le sikiana ou autre yugh, qui viendront rejoindre les lointains gaulois, thrace, illyrien et celtibère.

Des familles entières, voire des macrofamilles, seront bientôt rayées de la carte. C'est notamment le cas de la famille des langues iénisséennes, dont le khet, parlé en Sibérie centrale,

est le dernier survivant. C'est encore le cas, bien que dans une moindre mesure, de la macro-famille khoïsan, parlée en Afrique du Sud et en Tanzanie. Certaines aires géographiques sont plus touchées que d'autres. Ainsi des langues paléo-sibériennes (dites encore paléo-asiatiques, ou hyperboréennes, dispersées sur l'immense territoire sibérien, surtout au Kamtchatka, comprenant le chukchi et le giliak), des langues indiennes d'Amérique du Nord, des langues aborigènes d'Australie. Processus naturel inévitable, ou produit de civilisations irresponsables ? Les opinions divergent. « Language vitality and endangerment as aspect of language evolution », énonce Mufwene, en tête de chapitre (2001, chap. VIII), dépassionnant le débat. « Halte à la mort des langues » dit, à rebours, le titre d'un récent livre de Hagège, déjà cité. Le premier a une approche écologique des phénomènes linguistiques, qui s'exprime en termes adaptatifs, de compétition et sélection langagière, de coûts et profits pour le locuteur. La vie et la mort des langues sont le résultat d'une dynamique d'intégration des individus et des groupes dans le corps socio-économique environnant. Dans le même esprit, Georges Mounin (1992, pp. 156-158) relativise l'attachement des locuteurs à la survie de leur langue ainsi que la corrélation entre langue, vision du monde et identité ethnique et culturelle. Pour le second, nourri par l'hypothèse Sapir-Whorf, reprise par les ethnologues et un certain nombre de linguistes comme James Matisoff (1980) et Dixon (1980), la valeur historique et anthropologique des visions du monde portées par ces langues disparaît avec elles, en même temps que des informations précieuses sur notre patrimoine linguistique. Leur perte est donc irréparable. Cependant, les décisions de la communauté parlante sont seules pertinentes.

5. Le foisonnement linguistique : trois modes de classification

Si l'on cherche à remonter dans le temps – à l'échelle des linguistes, la profondeur maximale se situe au Néolithique, de -8 000 à -10 000 ans avant le présent - comment appréhender le foisonnement des langues ? Quelles règles dynamiques sous-tendent cette turbulence ? (Voir Annexes 1, Encadré 1 : « La diversité des langues »).

5.1. La classification généalogique

Les linguistes ont longtemps privilégié l'approche généalogique. C'est aussi la mieux documentée. La métaphore de la « naissance » a pour corollaire celles de « langues mères » et de « langues filles » - bien que la seule réalité observable soit un processus insensible de diversification et que seul le recul de l'histoire permette de poser, avec certaines marges d'incertitude, des « familles » de langues (comme la famille tibéto-birmanne), des « sous-familles » ou « branches » (comme la branche des langues slaves). La classification généalogique est fondée sur la statistique lexicale portant sur un vocabulaire de base, et sur l'identification de correspondances phonétiques régulières. Calquée sur l'évolutionnisme, cette méthodologie a donné lieu à d'importantes avancées au XIXe siècle et surtout au XXe siècle, à commencer par les domaines indo-européen et sémitique. Cependant, il montre aussi ses limites (Voir Annexes 2, Cartes couleur : « Les langues austronésiennes... » ; « Le Caucase : mosaïque linguistique... »). Certaines langues demeurent toujours inclassées, comme le japonais – dont on ignore s'il faut le rattacher à l'ensemble altaïque – ou sont des isolats, seuls représentants d'une famille et sans relation démontrée avec aucune autre, comme le sumérien ou aujourd'hui le basque ou le bouroushaski. On reviendra plus loin (sections 5.2 et 6) sur les limites de cette classification, et on s'arrêtera (section 7) sur l'exemple du proto-indo-européen.

5.2. La classification aréale

À ce modèle vertical et arborescent, se superpose et parfois s'oppose la classification horizontale dite aréale (par aires géographiques). C'est un concept venu de l'ethnologie, tandis que le premier est issu de la biologie. Il fait apparaître un mode d'appareillage de type sériel. Au sein de la diversité, les faits homologues observés ne s'expliquent pas nécessairement par une origine commune. Ils peuvent être simplement dus au voisinage et aux relations continues qui s'ensuivent. Les considérations géographiques ont l'avantage d'être moins abstraites que celles de type généalogique. En l'absence de données historiques, comme c'est souvent le cas, origine commune et contagion sont deux hypothèses également envisageables. D'où une révision de certaines reconstructions généalogiques. Ainsi, le modèle généalogique semble insuffisant et impropre à rendre compte de la véritable évolution historique de nombreux ensembles linguistiques, comme celui des langues bantoues, au sein du domaine négro-africain, ou du quetchua, au sein du domaine amérindien, ou du turc et du mongol, dans le domaine altaïque. Ainsi encore, l'unité du groupe ouralo-altaïque n'est plus reconnue, celle du chamito-sémitique fait l'objet de dissensions : famille ou union ? Évolution divergente ou convergente ? Voici plus d'un siècle, Johannes

Schmidt critiquait déjà la méthode comparative du modèle arborescent, pour proposer que les innovations dialectales se répandent comme des ondes (Wellen-Theorie). La théorie fut reprise dans les années 1930 par Nicolas S. Troubetzkoy (1939), qui voyait les langues se grouper en « maillons de chaîne », ou encore en portions d' « arc-en-ciel », idée adoptée plus tard par Roman Jakobson dans le domaine de la phonologie.

Cette approche commence aujourd'hui à porter ses fruits. Elle ne fait pas obstacle aux concepts de structure et de système (au sens saussurien où une langue est un système où tout se tient).

Bertil Malmberg (1990, p. 16) a analysé dans ce cadre le phénomène d'affaiblissement d'un système linguistique, que ce soit à la périphérie (géographique, sociale) ou par rencontre avec un autre système (bilinguisme). On a pu aussi évoquer l'idée de cellules aux membranes perméables réagissant les unes avec les autres, ou encore d'osmose (Mufwene 2001, p. 15).

Comme le formule Mounin (1992, p. 152), la « vieille tradition de l'impénétrabilité réciproque des grammaires » est à dépasser. C'est ainsi que l'on peut rendre compte, par exemple, de la curieuse unité du domaine balkanique (le bulgare et le grec notamment présentent des homologies). Cette méthode a aussi permis de mettre en évidence (Haudricourt, 1954), en Extrême-Orient, l'existence d'une vaste chaîne, le tibéto-birman et le chinois étant apparentés au miao (langue sino-tibétaine parlée en Chine du Sud, au Vietnam, au Laos et en Thaïlande) et à l'austro-asiatique (famille de quelque 150 langues, dont le khmer et le vietnamien), ceux-ci au thaï, ce dernier au malayo-polynésien (dit encore austronésien, famille comprenant notamment l'indonésien et les langues polynésiennes. Voir Annexes 2, carte « Les langues austronésiennes »). De même, de grandes homologies de traits grammaticaux et phonologiques entre les langues paléo-sibériennes et certaines langues avoisinantes (comme entre le chukchi et l'esquimo, ou entre le giliak et le coréen ou le japonais). Les affinités entre des groupes tels que l'ouralo-altaïque et le dravidien (sud de l'Inde) peuvent également s'expliquer en termes de communication prolongée et intime autant qu'en terme de familles (Voir Annexes 2, Schémas couleur : « Les langues ouraliennes »).

La théorie aréale révèle ses limites lorsque les ressemblances entre langues observées s'étendent sur d'immenses aires, parfois discontinues, ce qui laisse le choix entre de nombreuses hypothèses. Les contacts entre langues peuvent être à la source de la réduction de la diversité (par un processus de convergence), ou au contraire engendrer la diversité. On a tenté d'expliquer la multiplicité des langues de Mélanésie en les faisant dériver de mélanges assez récents entre les langues indonésiennes et les langues papoues, qui ont donné naissance à des variétés disparates d'indonésien pidginisé avec des substrats papous dans chaque groupe insulaire. En tout état de cause, la parenté doit aujourd'hui équilibrer les considérations sur divergence, convergence et mélange, sans privilégier aucun de ces processus.

5.3. La classification typologique

Aux côtés de la méthode généalogique et de l'approche aréale, un troisième mode de classification s'est constitué. C'est la méthode dite typologique. Elle range les langues en types en se fondant sur les similarités de structures linguistiques. On peut ainsi distinguer avec Hagège (1982, pp. 9-10) les langues à distinction verbo-nominale (comme les langues indo-européennes et sémitiques), à relative indistinction verbo-nominale (comme le nootka de l'île de Vancouver ou le santali d'Inde), ou mixtes de ce point de vue (comme les langues sino-tibétaines). Puis les langues à construction ergative, marquant formellement l'agent (comme le basque ou langues du Sud du Caucase), celles à construction accusative, marquant formellement le patient (comme le français ou le malais), ou encore des langues mixtes de ce point de vue (comme les langues océaniques ou australiennes). On distingue aussi, depuis le début du XIXe siècle, les langues essentiellement agglutinantes (suffixes mis bout à bout, comme dans les langues ouraliennes), flexionnelles (verbes conjugués, noms déclinés, comme en russe), isolantes (mots le plus souvent d'une seule syllabe, invariables et inanalysables, comme le chinois) ou encore les langues à très riche système consonantique (comme les langues caucasiennes) ou vocalique (comme les dialectes alémaniques et les langues polynésiennes).

Élaborée par les frères Schlegel, puis par Wilhelm von Humboldt, cette classification par types a considérablement évolué. Elle a montré qu'en certaines aires, une grande diversité généalogique s'accompagne d'un ensemble de traits typologiques communs : il en est ainsi dans la sous-région continentale de l'Asie du Sud-Est (Chine du Sud, Est de l'Inde), en Afrique

du Sud et de l'Est, sur la Côte Nord-Ouest et le Plateau limitrophe d'Amérique du Nord (qui comprend les plateaux des régions de Colombie et du Colorado et les hauts plateaux mexicains à l'est de la Sierra Madre). Mais force est de constater que familles de langues et types linguistiques coïncident rarement. Dans l'aire tibéto-birmane, la proximité généalogique ne va pas de pair avec la proximité typologique. Et la classification par types n'a pu éliminer l'arbitraire dans le nombre de critères typologiques retenus.

La recherche typologique a cependant apporté un enseignement fondamental : en dépit de la vaste pluralité des langues, les différences restent contenues dans certaines limites. Elles renvoient à des ressemblances de fond, qui désignent les facultés langagières d'*Homo sapiens sapiens*. Ces ressemblances ne sont pas plus interprétables, à l'heure actuelle, que l'expansion de l'art paléolithique dans les mêmes formes et quasi simultanément en Tanzanie, Afrique de l'Est, Afrique du Sud, Australie, Inde et Amérique ; pas plus interprétables que l'apparition de systèmes de numération à base 3, 4, 5, 10 ou 20 selon différents points du Globe - ou encore que l'invention des différents types d'écriture (Voir Annexes 2, « Photo de divers types d'écriture »).

6. La reconstruction : jusqu'où ?

Avec ce fond commun en tête, les généalogistes sont tentés de reconstruire des « langues-ancêtres », ou proto-langues. Des macrofamilles, ou superfamilles, voire comme Merritt Ruhlen (1997), une proto-langue mondiale (« proto-world »). Certains se risquent plus loin encore, en tentant de construire des modèles du « proto-sapiens » (de Grolier, 1988, p. 203).

Ainsi, dès 1903, Holger Pedersen, tenta de reconstituer une langue-mère commune dite nostratique, aujourd'hui toujours d'actualité mais très discutée. Vladislav M. Illitch-Svitytch (1976) et Aaron B. Dolgopolski (1986) reprirent le flambeau vers 1960, la faisant remonter au début du Néolithique. Six familles en dériveraient, dont l'indo-européen et le sémitique (Jucquois, 1996, pp. 152-154). Un autre unificateur, célèbre et moins contesté, est Joseph H. Greenberg, (2003) qui a notamment reconstruit la superfamille eurasiatique, au sein de laquelle se rangerait l'indo-européen. Mais l'unificateur Ruhlen, en collaboration avec John Bengston, a proposé, quant à lui, pour asseoir le proto-world, un ensemble de 27 racines mondiales, mots qui se retrouvent dans les 12 macrofamilles et qui regroupent, selon lui, toutes les langues du monde pour désigner à peu près le même référent.

Cette hypothèse est loin d'être admise par la communauté des linguistes. Trois arguments notamment paraissent la réfuter. Le premier est d'ordre extralinguistique. Les modèles mathématiques de type probabiliste émanant de certains chercheurs indiquent en effet qu'en suivant la méthode de Ruhlen, les chances de trouver des cognats dans toutes les langues sont de 100 % (Boë, cité par Coupé, 2005, p. 173). Le second est d'ordre linguistique. Au-delà de ce qu'on appelle le premier degré, c'est-à-dire dès qu'on aborde le stade de la superfamille, les données linguistiques, mais aussi culturelles et historiques se raréfient. Selon la formule de Guy Jucquois (1996, pp. 152-154), « on glisse progressivement des « correspondances » aux « ressemblances » ». La méthode comparative poussée à ses extrémités et vers ces glissements conduit à travailler « sous le niveau du bruit », c'est-à-dire vers une indétermination foncière entre ressemblances accidentelles et ressemblances dues à une histoire commune entre les langues (Boë cité par Coupé, 2005, p. 173). Le troisième argument est encore d'ordre linguistique. La plupart des chercheurs estiment aujourd'hui que l'applicabilité de la méthode comparative devient nulle après environ 7 000 à 10 000 ans de séparation. S'inspirant de la datation de restes organiques par le carbone 14, Morris Swadesh (1952, pp. 452-453) avait inauguré la méthode lexicostatistique de datation des langues dite glottochronologique. Elle est utilisée pour calculer la durée de séparation entre des langues apparentées et se fonde sur le taux de remplacement du vocabulaire de base par périodes de 1000 ans. À sa suite, Hagège, ainsi que Georges André Haudricourt, et parallèlement Lionel Bender ont proposé une théorie cyclique de l'évolution des langues. Celles-ci évolueraient, en transformant leurs lexiques et leurs structures à un rythme relativement constant, tant et si bien qu'au bout de la longue période indiquée, elles se seraient entièrement renouvelées. Que l'on songe simplement à l'exemple phonétique du mot signifiant « langue » dans le domaine indo-européen. « Le prototype que l'on reconstruit, écrit Martinet (1986, pp. 128-130) a la forme « d₁ng₁hweH₂ ». Par une série de transformations régulières à travers les siècles, l'auteur montre qu'il a pu aboutir à des formes aussi contrastées que le latin *lingua*, l'anglais *tongue*, le

russe *jazyk* et le tokharien *käntwa*. Par ces arguments, l'horizon des linguistes se trouve en tout cas sérieusement raccourci. À supposer même que l'on puisse démontrer un jour l'existence d'une langue mère de toutes les langues, cette langue reconstruite ne saurait se confondre avec le concept de « langue originelle ». Antonio Tovar (1954) décrivait voilà plus d'un demi-siècle, l'illusion de ce qu'il appelait « le vieux fantôme de l'arbre généalogique ».

Dans la reconstruction à l'infini, dans certaines tentatives de macro-comparaisons se profile en réalité la recherche de la langue adamique, d'avant Babel. Il en est de même de la « quête d'une langue parfaite » évoquée par Umberto Eco (1992). Vaine quête du Graal... Des disciplines comme les mathématiques (théorie des probabilités), l'archéologie, la paléontologie et l'éthologie viennent ébranler l'ancienne utopie. Il n'est pas à exclure en effet que, dans un lointain passé, plusieurs langues mères se soient succédées ou encore aient coexisté, et qu'une seule ait survécu par la suite (Coupé, 2005, p. 186). De nombreux scénarios sont envisageables.

7. Le proto-indo-européen : une marge d'indétermination

Pour redescendre plus modestement de l'arbre généalogique au niveau des familles du premier degré, la prudence s'impose encore. Il est vraisemblable que les similitudes identifiées au cours des tentatives de reconstruction soient davantage le fait de la procédure utilisée, de la succession des inférences, que de la nature réelle de la proto-langue qu'on pense avoir caractérisée. Comme l'énonce Jucquois (1996, p. 156) :

La comparaison génétique aboutit ainsi à une illusion partielle puisqu'elle focalise en un temps originel les résultats de la comparaison supposés être synchrones par la simplicité qui résulte pourtant du simple aboutissement de la procédure de comparaison.

Ne serait-ce que pour la seule famille indo-européenne, la plus anciennement documentée historiquement, l'incertitude demeure entière. Les conditions initiales étaient-elles homogènes ou hétérogènes ? Déterminer une famille de langues est une chose (cela relève de la taxinomie), chercher une langue ancêtre en est une autre (on passe au domaine de la reconstitution). Faut-il donc poser une langue mère unique, comme le croient certains ? Faut-il plutôt voir, à l'origine, un amas linguistique qui a donné naissance au groupe de langues ayant par la suite subi un phénomène de convergence, sous le double effet du contact et du temps ? C'est la thèse de Troubetzkoy (1939) reprise par Tovar et d'autres. Ou bien peut-on imaginer une forme originellement dialectale, donc multiple ? À ces incertitudes ontologiques s'ajoutent les difficultés de localisation spatiale du foyer originel supposé : foyer anatolien selon l'archéologue Colin Renfrew (1989). C'est la thèse dominante actuellement, en raison de données archéologiques et génétiques convergentes, mais l'auteur reconnaît lui-même que son hypothèse ne peut tenir que si elle n'entre pas en conflit avec les données de la comparaison des langues), russe du sud-est selon l'archéologue Maria Gimbutas ainsi que Martinet (1986), boréal en Eurasie septentrionale selon Nicolas D. Andreïev, et foyers multiples entre la Mer du Nord et la Mer Caspienne selon Troubetzkoy. Sans parler des incertitudes sur la date d'émergence du proto-indo-européen : elle varie, selon les auteurs de -8 500 à -6000 avant le présent. « Ceci demeure, quoi qu'il en soit, une pure supposition. Il aurait tout aussi bien pu être parlé il y a 15 000 ans », renchérit, pour sa part, Alain Peyraube (2005, p. 344).

8. Monogénèse et polygénèse : le débat n'est pas tranché

Dans l'état actuel des connaissances, il est impossible de se prononcer sur la monogénèse ou la polygénèse. Les avancées de la génétique et de la paléoanthropologie tendent à privilégier aujourd'hui l'hypothèse de l'évolution unirégionale, dite « Out of Africa », ou « Arche de Noé ». Celle-ci propose que l'*Homo sapiens sapiens* actuel serait originaire d'Afrique de l'Est et issu d'une seule population d'*Erectus* (en ayant partout supplanté ce dernier). L'autre hypothèse, multirégionale, dite « Candélabre » préconise qu'une population d'*Erectus* a évolué de façon parallèle et indépendante pour donner naissance à des groupes de « *Sapiens régionaux* » sur toute la planète. La majorité des données génétiques va ainsi dans le sens d'une origine récente et africaine de notre espèce. Toutefois, l'étude de la diversité génétique de certaines régions du génome contredit cette proposition (Quintana-Murci & Hombert, 2005, pp. 312-313, 315, 317). En réalité, les résultats obtenus dépendent des marqueurs génétiques choisis, et par conséquent, la reconstruction génétique est liée à la procédure utilisée. Quoi qu'il en soit, si la thèse unirégionale l'emportait, elle ne serait pas pour autant incompatible avec la polygénèse linguistique, comme on l'a vu plus haut (section 6).

Ce n'est, encore une fois, que de l'étroite collaboration entre les préhistoriens, les archéologues, les généticiens et les linguistes que pourront surgir des éléments de réponse.

9. L'hypothèse de l'aire ancestrale

Il existe, en revanche, un certain consensus sur l'hypothèse plus modeste, de l' « aire ancestrale », énoncée dans l'entre-deux guerres par Edward Sapir. Si l'on veut identifier le foyer originel d'une famille linguistique, disait-il, la zone de plus grande ancienneté historique correspond à celle de la plus grande divergence. Ainsi, c'est en Mélanésie, entre l'Archipel Bismarck et les Nouvelles Hébrides que se trouve la plus grande diversité des langues austronésiennes. Cela semblerait indiquer que le proto-indonésien était parlé là voici des millénaires. Or, l'archéologie vient à l'appui de cette thèse, car c'est aussi en Mélanésie qu'ont été trouvés les assemblages d'objets façonnés les plus anciens de cette partie du monde. En Afrique, c'est la région située au Nord de la zone bantoue et au Sud du Sahara (nord de la frontière Cameroun-Nigéria) qui est la plus diversifiée linguistiquement. On suppose donc, en suivant ce raisonnement, que c'était là le foyer originel des Proto-Bantous. Or, cela semble être conforté par les données génétiques fondées sur l'analyse de certaines protéines. Dans le domaine amérindien du Nord, la plus grande diversité des langues subarctiques conduit à l'hypothèse d'une migration athabasque à partir de cette aire.

10. Gènes et langues : corrélation ?

On a vu que la génétique, n'échappe pas, quant aux hypothèses de l'origine uni- ou multi-régionale des *Sapiens*, à des problèmes liés à la procédure utilisée. De la même façon, les corrélations entre langues, populations et cultures à travers le monde sont difficiles à tracer. Les contrées notoires pour leur diversité linguistique ne recouvrent pas d'hétérogénéité ethnologique qu'on serait en droit d'attendre. Des exemples contraires comme les Bochimans d'Afrique du Sud ou les Berbères d'Afrique du Nord ne sont que des exceptions. Les récentes découvertes de l'école de Luigi Cavalli-Sforza (1996) concernant la correspondance entre certains groupes linguistiques et humains témoignent de la complexité des phénomènes. Non seulement gènes et langues ne coïncident pas toujours, mais les résultats, là encore, divergent en fonction des marqueurs génétiques choisis. Le chromosome Y (révélateurs de l'histoire des hommes) montre une corrélation plus forte avec la linguistique que l'ADN mitochondrial (révélateur de l'histoire des femmes). Or, selon le premier marqueur, ainsi que les marqueurs classiques, les Basques, dont la langue est un isolat, sont aussi isolés génétiquement. En revanche, selon le second marqueur, les populations basque et indo-européenne sont très homogènes génétiquement. Histoire migratrice ou histoire culturelle différentes ? On l'ignore. Il en est de même pour les populations et les langues sames. Selon des analyses raffinées d'ADN mitochondrial, pas de coïncidence entre gènes et langues : les gènes sont proches de ceux portés par les populations de langues indo-européennes, tandis que les langues appartiennent à la famille ouralienne. Dans le sens inverse, les Tunisiens, Egyptiens et Ethiopiens, pourtant géographiquement séparés par plusieurs milliers de kilomètres, appartiennent à la même macro-famille linguistique afro-asiatique (celle-ci n'est pas toutefois reconnue par tous) et se révèlent très proches génétiquement (Quintana-Murci & Hombert, 2005, pp. 324-325). La prudence s'impose donc en ces matières. D'autant plus que l'équation langue=race - et toute l'idéologie mortifère relative aux supériorités et infériorités supposées - rôde désormais comme un spectre. Quoi qu'il en soit, les avancées de l'école Cavalli-Sforza, aussi intéressantes que prometteuses, plaident autant pour un apparemment aéal ancien que pour un ancêtre commun.

11. Homogénéité et hétérogénéité linguistique

Chaque idiome recèle ses clefs de la diversité linguistique. La parfaite homogénéité d'une langue est une vue de l'esprit. En réalité, l'hétérogénéité et la variation caractérisent l'être même de la langue - comme l'ont montré les travaux importants de Martinet, d'Uriel Weinreich (1968), de Hagège (1993), William Labov (1994) et Mufwene. Pour le premier, structuraliste et fonctionnaliste, et inventeur du concept de « synchronie dynamique », variation et changement linguistique se font *dans* et *par* la structure (1964, pp. 521-532) ; « une langue change parce qu'elle fonctionne » résume-t-il (1990, p. 155). Ce en quoi il est très proche de la biologie évolutionniste non fixiste, qui lit les structures et les fonctions comme résultat d'une histoire, et encore en devenir. Le second consacre un volume entier (1993) à la dynamique des langues, mettant à l'avant-scène l'activité consciente et inconsciente des locuteurs-constructeurs.

Pour le dernier, plutôt ancré dans la sociolinguistique (2001, p. 24) :

The agency of change lies definitely within the behavior of individual speakers, and causation partly in the mutual accommodations they make to each other while they are more intent on communicating effectively than on preserving idiolectal, dialectal or language boundaries.

Les causes de l'hétérogénéité sont des plus diverses : paramètres sociaux, de sexe, d'âge, d'habitat, d'ethnie, de discours, de style... Aujourd'hui, grâce à la paléogéographie, on peut ajouter un nouveau facteur de variation : la taille du groupe de locuteurs. Celle-ci conditionnerait en effet les structures linguistiques mêmes d'une langue donnée (Bocquet-Appel, cité par Coupé, 2005, p. 190). L'hétérogénéité est à l'oeuvre jusque chez le locuteur, chacun de nous n'en finissant pas de réorganiser notre langue. Dans tout processus de communication, opère un double mouvement d'homogénéisation et d'hétérogénéisation. Des zones de stabilité relative ainsi que d'instabilité sont détectables en synchronie.

12. Conclusion

La question des origines n'est désormais plus taboue pour les linguistes. On peut s'en réjouir car cette problématique est abordée par bien d'autres sciences, naturelles et humaines. Loin de constituer une vaine quête du Graal, comme le serait celle d'une « langue parfaite », elle s'avère féconde, dans notre champ, lorsqu'elle s'inscrit dans une démarche pluridisciplinaire. Néanmoins, vu la complexité des données, la prudence méthodologique doit constamment guider nos pas. Ceci d'autant plus que toute remontée aux sources premières n'est pas exempte de passion, traversée par les mythologies, et les idéologies qui nous environnent. Qu'il s'agisse de l'origine de l'univers, de la vie, de l'homme, ou du langage et des langues, elle est sporadiquement sous-jacente, et il ne manque pas de chercheurs (à l'instar de l'astrophysicien Hubert Reeves, du cosmologue Jayant Vishnu Narlikhar, du chimiste Jacques Reisse et du paléoanthropologue Yves Coppens) pour stigmatiser ses intrusions dans la science. Dans le domaine linguistique, on observe que les théories sur le monogénisme et le polygénisme ont toutes deux été des vecteurs d'idéologies douteuses, conduisant quelquefois au racisme.

Tour à tour désacralisée et libérée des tabous, cette délicate question peut prendre toute sa place comme l'une des clés explicatives de la diversité linguistique. Toutefois, notre dispositif cérébral, les conditions d'évolution des *Sapiens* ainsi que leur histoire migratrice ne sauraient rendre compte à eux seuls, bien entendu, de la diversité observée. L'usage que font de la langue les locuteurs en perpétuelle dynamique adaptative, et en incessante situation de contact, ainsi que la remise en question progressive des valeurs conventionnelles constitutives des langues, en tant qu'institutions sociales, demeurent un très puissant moteur du changement linguistique et de la diversité.

L'attention portée à la pluralité linguistique, a en outre l'avantage de mettre en relief, sur fond d'universalisme (l'unité écologique et psycho-physiologique du genre humain, et le fait que l'on puisse traduire d'une langue à l'autre) un relativisme linguistique fondamental. Faire voir ce que, selon la formule de Jakobson reprise par Hagège, les langues « obligent » à dire ou « empêchent de dire », c'est aussi prendre conscience que les structurations spécifiques des idiomes imposent aux sujets parlants des représentations diversifiées du monde qu'il importe d'analyser dans leur richesse, non de réduire à l'unique.

Références bibliographiques

- Bickerton (D.). 1981 : *Roots of Language*. Ann Arbor : Karoma.
- Boë (L.-J.), Bessière (P.) & Vallee (N.). 2003. « When Ruhlen's « mother tongue » theory meets the null hypothesis ». in : *Proceedings of the XVth International Congress of Phonetic Sciences*. Barcelone.
- Chaudenson (R.). 2000. « Peut-on décrire un créole sans référence à sa genèse ? Pronoms et adjectifs dans les créoles français ». in : *Langages*. Paris : Larousse.
- Cavalli-Sforza (L.). 1996. *Gènes, peuples et langues*. Paris : Odile Jacob.
- Coupé (C.). 2005. « L'impossible quête de la langue mère ». in : *Aux origines des langues et du langage*. Paris : Fayard, pp. 161-195.
- Dixon (R.-M.-W.). 1980. « The Endangered Languages of Australia, Indonesia and Oceania ». in : Robin & Uhlenbeck (eds.). *Endangered Languages*. Oxford, New York : Berg, pp. 229-255.
- Dixon (R.-M.-W.). 1997. *The Rise and Fall of Languages*. Cambridge : Cambridge University-Press.
- Dolgopolski (A.-M.). 1986. « A Probabilistic Hypothesis Concerning the Oldest Relationships Among the Languages Families in the Northern Eurasia ». in : Shevoroskin & Markey (eds.). *Typology, Relationship and Time*. Ann Arbor : Karoma.
- Eco (U.). 1992 : « La quête d'une langue parfaite dans l'histoire de la culture européenne ». [Leçon inaugurale au Collège de France, le 2 octobre 1992].
- Greenberg (J.-H.). 2003. *Les langues indo-européennes et la famille eurasiatique*. Paris : Belin [trad. fr.].
- Grolhier (E. de). 1988. « Aux origines du langage ». in : Brenot (P.), (dir.). *Les origines*. Paris : L'Harmattan, pp. 187-222.
- Hagège (C.). 1982. *La structure des langues*. Paris : P.U.F., Coll. « Que sais-je ? ».
- Hagège (C.). 1985. *L'homme de paroles*. Paris : Fayard.
- Hagège (C.). 1993. *The Language Builder : an Essay on the Human Signature in Linguistic Morphogenesis*. Amsterdam : Benjamins.
- Hagège (C.). 2000. *Halte à la mort des langues*. Paris : Odile Jacob.
- Hagège (C.). 2003. « Le langage et les langues entre le biologique et le social ». in : Changeux (J.-P.), (dir.). *Gènes et culture*. Paris : Odile Jacob, pp. 153-172.
- Haudricourt (G.-A.). 1954. « Comment reconstruire le chinois archaïque ». in : *Word*, 10.
- Hazaël-Massieux (M.-C.). 2005. « Au sujet de la définition des créoles ». in : *La Linguistique*, 41, 1 : *Les créoles*, pp. 3-17.
- Hombert (J.-M.). 2005. « Conclusion ». in : Hombert (J.-M.), (dir.). *Aux origines du langage et des langues*. Paris : Odile Jacob, pp. 462-465.
- Hublin (J.-J.). 2003. « Évolution des hominidés et origine du langage ». in : Changeux (J.-P.), (dir.). *Gènes et culture*. Paris : Odile Jacob, pp.17-33.
- Hublin (J.-J.). 2005. « La langue des premiers hommes ». in : Hombert (J.-M.), (dir.). *Aux origines des langues et du langage*. Paris : Fayard, pp. 102-117.
- Illitch-Svitytch (V.-M). 1976 : *Opyt sravnenija nostratičeskix jazykov* [Étude comparative des langues nostratiques]. Moscou : Nauka.
- Jucquois (G.). 1996. « Le nostratique ». in : *La Linguistique*, 32, 2. Paris : P.U.F., pp.152-154.
- Labov (W.). 1994. *Principles of Linguistic Change : Internal Factors*. Oxford : Blackwell.
- Mandel (J.-L.). 2003. « Maladies monogénétiques et troubles de la cognition ». in : Changeux (J.-P.), (dir.). *Gènes et culture*. Paris : Odile Jacob, pp. 35-56.
- Malmberg (B.). 1990. « Linguistique aréale et recherches comparatives ». in : *Mémoire de la SLP*. Nouvelle Série, tome I. Paris : Klincksieck, p. 16.
- Martinet (A.). 1964. « Structural variation in language ». in : *Proceedings of the 9th International Congress of Linguists*, pp.521-532.
- Martinet (A.). 1968. « Préface ». in : Martinet (A.), (dir.). *Le Langage. Encyclopédie de la Pléiade*. Paris : Gallimard, pp. 6-12.
- Martinet (A.). 1986. *Des steppes aux océans. L'indo-européen et les « Indo-Européens »*. Paris : Payot.
- Martinet (A.). 1990. « Changements linguistiques ou fonctionnement du langage ? ». in : *La Linguistique*, 26, 2, pp. 153-158.
- Matisoff (J.A.). 1980. « Endangered Languages of Mainland Southeast Asia ». in : Robin & Uhlenbeck (eds.). *Endangered Languages*. Oxford, New York : Berg, pp. 189-228.

- Meillet (A.). 1965. « Le problème de la parenté des langues ». in : Meillet (A.). *Linguistique historique et linguistique générale*. Paris : Honoré Champion, pp. 76-109 [rééd.].
- Mounin (G.). 1992. « Sur la mort des langues ». in : *La Linguistique*, 28, 2, pp. 149-158.
- Mufwene (S.-S.). 1986. « Les langues créoles peuvent-elles être définies sans allusion à leur histoire ? ». in : *Études créoles*, 9, 1, pp. 135-150.
- Mufwene (S.-S.). 2001. *The Ecology of Language Evolution*. Cambridge : Cambridge University Press.
- Nespoulous (J.-L.). 2005. « La désintégration du langage dans le cerveau lésé de l'aphasique » in : Hombert (J.-M.), (dir.). *Aux origines du langage et des langues*. Paris : Odile Jacob, pp. 428-449.
- Nettle (D.). 2002. *Linguistic Diversity*. Oxford : Oxford University Press.
- Peyraube (A.). 2005. « Comment classer les langues ». in : Hombert (J.-M.), (dir.). *Aux origines du langage et des langues*. Paris : Odile Jacob, pp. 328-345.
- Peyraube (A.) 2005. « L'homme aux 6000 langues ». in : Hombert (J.-M.), (dir.). *Aux origines du langage et des langues*. Paris : Odile Jacob, pp. 450-461.
- Quintana-Murci (L.) & Hombert (J.-M.). 2005. « Gènes et langues : une évolution parallèle ? ». in : Hombert (J.-M.), (dir.). *Aux origines du langage et des langues*. Paris : Odile Jacob, pp. 308-327.
- Renfrew (C.). 1989. *Archaeology and Language. The Puzzle of Indo-European Origin*. London : Penguin Press.
- Ruhlen (M.). 1997. *L'origine des langues. Sur les traces de la langue-mère*. Paris : Belin [trad. fr.].
- Swadesh (M.). 1952. « Lexicostatistic Dating of Prehistoric Ethnic Contacts ». in : *Proceedings of the American Philosophical Society*, 96, pp 452-463.
- Szulmajster-Celnikier (A.). 1997. « À propos de *Langue et idéologie : les mots étrangers en hébreu moderne* - de Michel Masson ». in : *La Linguistique*, 33, 1, pp. 141-145.
- Tovar (A.). 1954. « Linguistics and Prehistory ». in : *Word*, 10, p. 348.
- Trubetzkoy (N.-S.). 1939. « Gedanken über das Indogermanenproblem ». in : *Acta Linguistica*, I, 2, pp 81-89. Copenhague.
- Weinreich (U.), Labov (W.) & Herzog (M.). 1968. « Empirical foundations for a theory of language change ». in : Lehman & Malkiel (eds.). *Directions for Historical Linguistics*. Austin : University of Texas Press.

Annexes 1

Encadré 1 : La diversité des langues

La diversité des langues est inégalement répartie sur la surface du Globe. Certaines régions sont connues pour leur foisonnement : Amérique du Sud, Nouvelle Guinée, province du Plateau au Nigéria, province d'Oaxaca au Mexique, certains territoires du Nord-Est de l'Inde. En Afrique, ce sont les langues bantoues, appartenant à la famille Niger-Congo, qui dominent par leur nombre (900) et celui de leurs locuteurs (160 millions). En revanche, les langues d'Europe, et singulièrement d'Europe occidentale, présentent un faible degré de diversité. C'est le résultat de conditions historiques particulières. D'une part, les contacts de populations et de leurs langues ont fait de l'Europe une aire caractérisée par des emprunts multiples et multidirectionnels. D'autre part, l'édification de normes linguistiques communes y a été très précoce et très générale. C'est en Europe que la réduction, voire l'élimination des variations dialectales a été le plus efficace. Mais on trouve aussi des exemples de larges groupes à l'homogénéité frappante sans intervention du politique. C'est le cas de la famille australienne aborigène, qui connaît un continuum dialectal étendu sur des distances considérables. À mi-chemin, le groupe slave se propage en maillons de chaîne. Divergence et convergence semblent y avoir joué un rôle équilibré.

C'est la famille indo-européenne qui compte le plus grand nombre de locuteurs. Elle est suivie par la famille sino-tibétaine, dans laquelle le chinois occupe une place centrale par sa masse numérique. Sa forme la plus répandue, le mandarin, détient de ce point de vue un record mondial (610 millions de locuteurs natifs). À l'autre extrémité du spectre, certaines langues en péril comme les langues paléo-sibériennes, ne comptent plus que quelques centaines de locuteurs (400 pour le youkaghir, langue de Sibérie arctique, à l'Est de la Léna ; une centaine seulement pour le khet, dans le bassin du Ienisséi).

En termes d'espace, certaines familles et groupes de langues s'étendent sur des aires impressionnantes : il en est ainsi de la famille austronésienne (Voir carte « Les langues austronésiennes »), de la famille indo-européenne, ou du groupe samoyède (langues ouraliennes). On voit aussi des langues peu parlées mais dispersées sur de vastes territoires : telles le yidiche, le romani (ou tzigane), ou encore l'arménien. Parfois, à l'inverse, un pullulement de langues se concentre dans un périmètre très réduit. Ainsi l'ensemble caucasien est confiné et morcelé dans les monts du Caucase, région nommée « mont des langues » par les géographes arabes du Xe siècle (Voir carte « Le Caucase : mosaïque linguistique »). C'est parfois pour faire face à une diversité ressentie comme trop forte que s'élaborent des *lingua franca* (au sens où l'anglais est devenu la langue des scientifiques), des *koine*, dialectes véhiculaires, des langues supralocales et fédératrices, comme le swahili en Afrique de l'Est, ou encore des pidgins comme le bichelamar (beach-la-mer) aux Nouvelles Hébrides et en Australie.

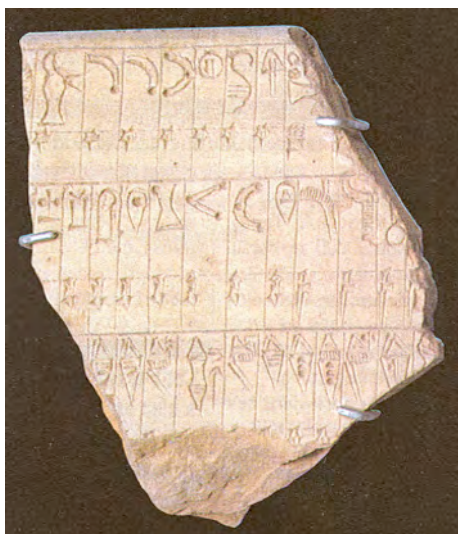
Encadré 2 : Quatre domaines de divergence

Les principaux domaines de divergence entre les langues sont par ordre croissant de source de divergence la syntaxe, la phonologie, le lexique et la morphologie. Sur une base commune d'organisation et d'expression de contenus plus ou moins universels propre à *Homo sapiens sapiens*, les langues divergent d'abord par leur structuration syntaxique. Les divergences sont plus fortes en phonologie, dans les limites imposées par la constitution des organes de la parole et de l'audition. Sur le plan lexical, les différences se manifestent non seulement entre les langues, mais aussi d'un individu à l'autre. C'est la morphologie qui constitue le champ de plus forte différenciation et partant, le domaine où les coïncidences interlinguistiques se font rares.

Les neuropsychologues, grâce à la technique d'imagerie cérébrale, viennent aujourd'hui conforter la réalité de l'architecture cérébrale postulée par les linguistes : la phonétique, la phonologie, le lexique et la syntaxe renvoient bel et bien à des représentations et des processus distincts dans le cerveau humain.

Annexes 2

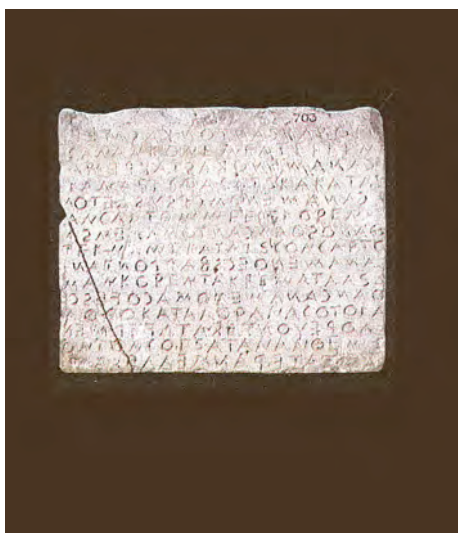
Image 1 : 7 photos de divers types d'écriture [1a à 1g]



(1a) Inscription assyrienne, 680 av. J.-C.



(1b) Inscription étrusque, Ve siècle avant J.-C.



(1c) Inscription dorienne de Crète, époque archaïque.



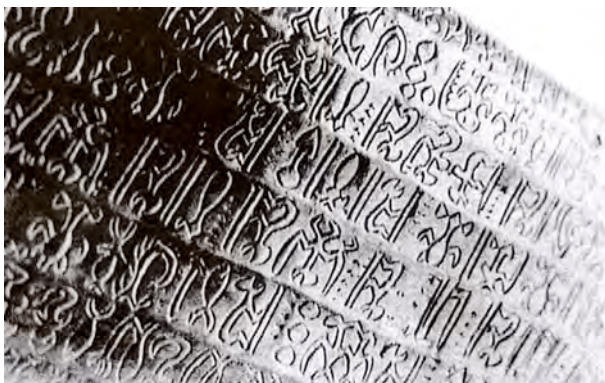
(1d) Hiéroglyphes hittites de la langue Iuwi, Xe-VIe siècle av. J.-C.
Écriture rarissime de type « boustrophédon »
(lecture horizontale en zigzag).



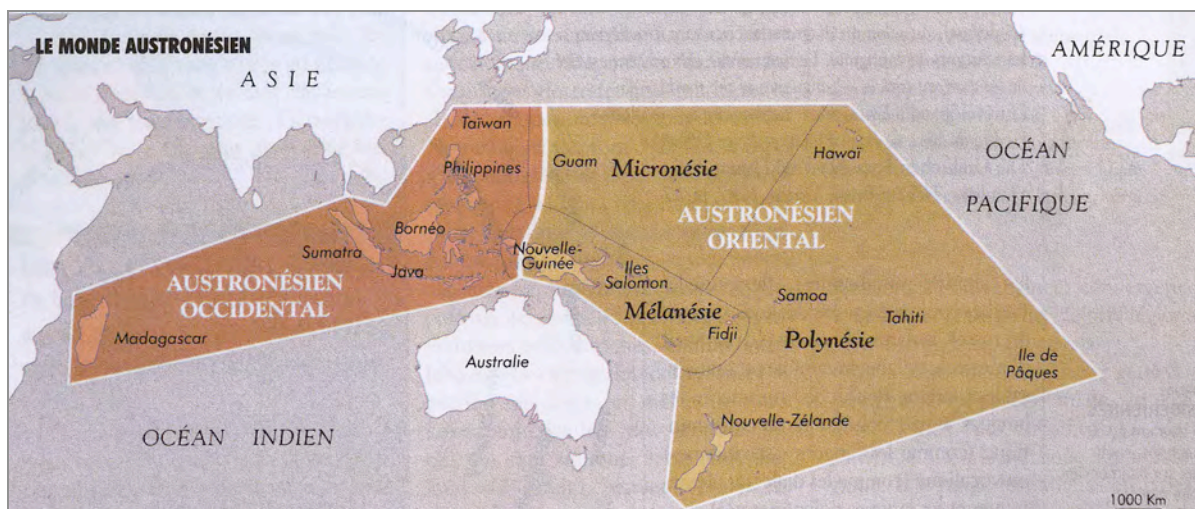
(1e) Inscription sumérienne, 2000 ans av. J.-C.



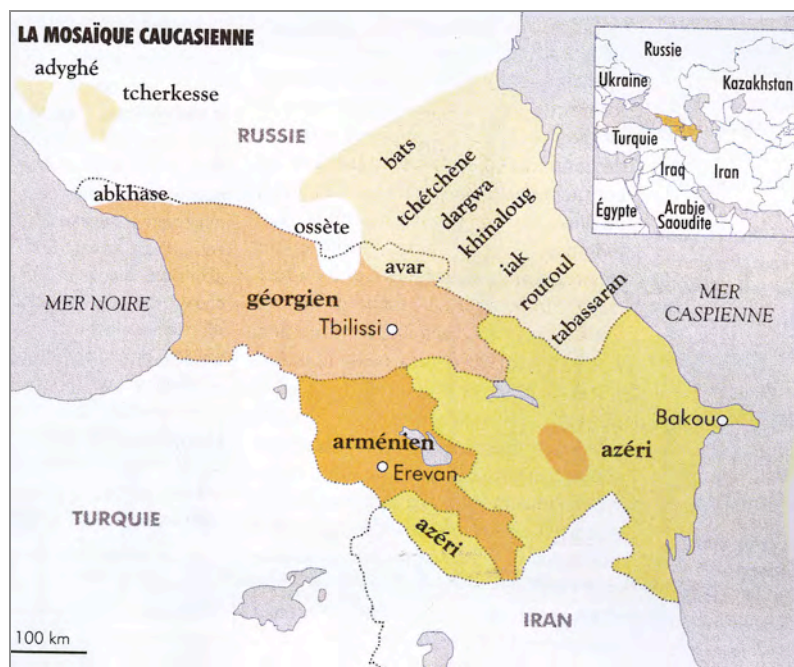
(1f) Inscription sumérienne, 2550 ans av. J.-C.



(1g) (1h) Hiéroglyphes de l'Île de Pâques (ici « Le bâton de Santiago »), en exploitation jusqu'au XIXe siècle, dont le déchiffrage est très récent. Écriture unique au monde de type « boustrophédon à inversion alternée » (lecture horizontale en zigzag, tête bêche d'une ligne à l'autre).



Carte 1 : Les langues austronésiennes : vaste extension d'environ 23 600 km. Pour expliquer la parenté entre ces langues, on en est réduit à échafauder des hypothèses.
 (Source : article « Languages of the world », Encyclopaedia Britannica).



Carte 2 : Le Caucase : mosaïque linguistique sur un espace réduit d'environ 1 200 X 600 km, et sur un relief très accidenté. On y distingue des langues des familles caucasienne, indo-européenne et altaïque.
 (Source : Hagège (C.). 1992. Le souffle de la langue. Paris : Odile Jacob, carte 3).

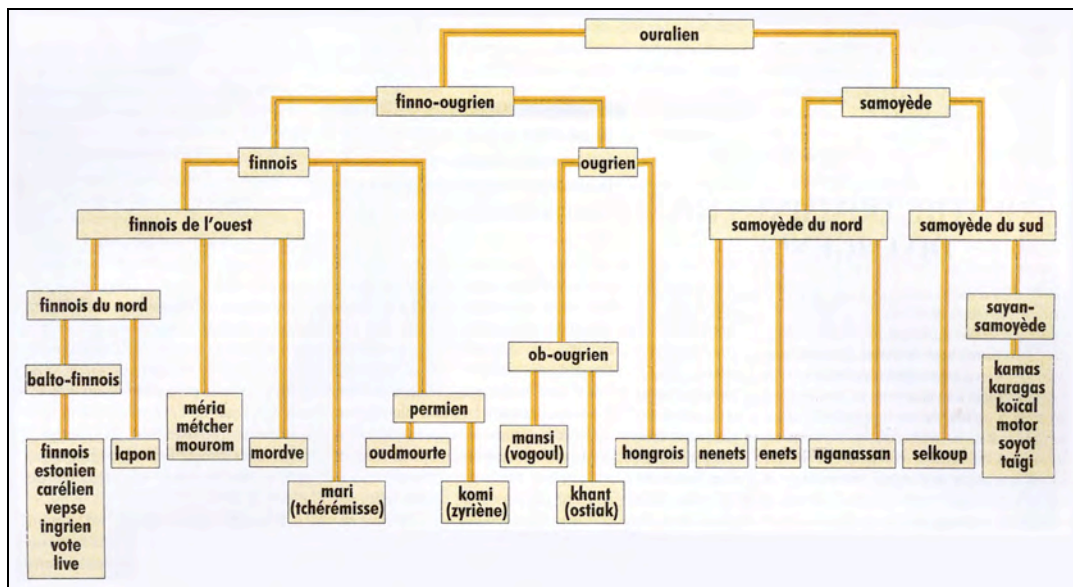


Schéma 1 : Les langues ouraliennes selon un diagramme arborescent, de type généalogique.
(Source : article « Languages of the world », Encyclopaedia Britannica).

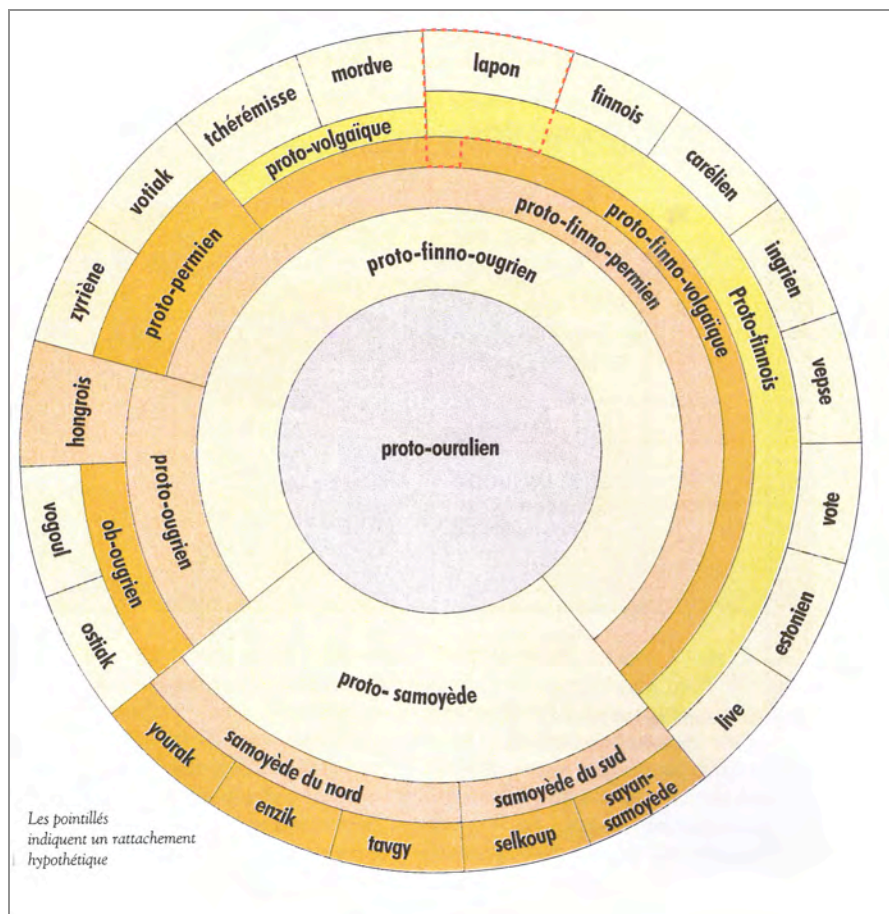


Schéma 2 : Les langues ouraliennes selon un diagramme circulaire, privilégiant les relations de contiguïté.
(Source : Hajdù (P.) & Dometlos (P.). 1987. Die Uralischen Sprachen und Literaturen. Hamburg: Helmut Buske Verlag, p. 312).