

2

Chapitre 5

LANGUES ET GÈNES

On parle aujourd'hui à peu près cinq mille langues différentes dans le monde. Quelques-unes sont utilisées par des centaines de millions de personnes, mais la plupart sont moins largement diffusées; certaines sont même limitées à une centaine d'individus, et il y a de fortes chances pour qu'elles disparaissent complètement d'ici quelques générations. Nous en connaissons beaucoup qui ont déjà disparu ou sont en train de le faire.

Il n'est pas nécessaire d'être linguiste pour s'apercevoir que certaines langues sont fort proches les unes des autres. Cela est vrai par exemple pour le français et pour l'italien, ma langue maternelle, et peut-être encore plus pour l'espagnol et l'italien. Si je vais en Espagne ou au Portugal, je me débrouille sans trop de difficultés, sauf lorsque je tombe sur deux mots qui sont identiques dans les deux langues ou qui se ressemblent beaucoup, mais qui ont une autre signification, comme *burro*, qui signifie «beurre» en italien et «âne» en espagnol. On appelle ces mots des faux amis, mais heureusement ils ne sont pas très nombreux. La raison

de la ressemblance entre les langues italienne, française, espagnole, roumaine, etc., est simple : toutes ces langues dérivent d'une source commune, le latin. Une ressemblance identique unit les langues d'origine germanique : scandinave, allemande, hollandaise, flamande et anglaise. Dans l'est de l'Europe, le groupe de langues slaves montre aussi des ressemblances très précises. Au XVIII^e siècle déjà, on s'était aperçu que la langue indienne classique, le sanscrit, ressemblait aux langues anciennes européennes. Comme cela arrive pour toute découverte ou invention, il y avait, voici quelques siècles, des pionniers de cette découverte.

Le XVIII^e siècle a vu naître de nombreux travaux linguistiques qui ont permis de mieux comprendre les parentés d'un groupe de langues qu'on appelle, à juste titre, la famille des langues indo-européennes. Alors on a commencé de reconnaître l'existence de beaucoup d'autres familles (certains linguistes préfèrent les appeler *phyla*, ou *phylum* au singulier). Comme les taxonomistes des plantes et des animaux, les biologistes ont reconstruit des arbres que les linguistes nomment « génétiques » – l'équivalent de phylogénétique en biologie. Mais les linguistes ont rencontré des difficultés à remonter plus haut que les familles de langues. Même si plusieurs d'entre eux ont cherché à établir des parentés entre les familles, il a jusqu'ici été impossible de s'accorder sur un seul arbre qui comprenne toutes les familles existantes. En effet, de nombreux linguistes pensent qu'on ne peut même pas répondre à la question de l'unicité ou de la multiplicité de l'origine des langues modernes. La difficulté réside dans la vitesse d'évolution des langues.

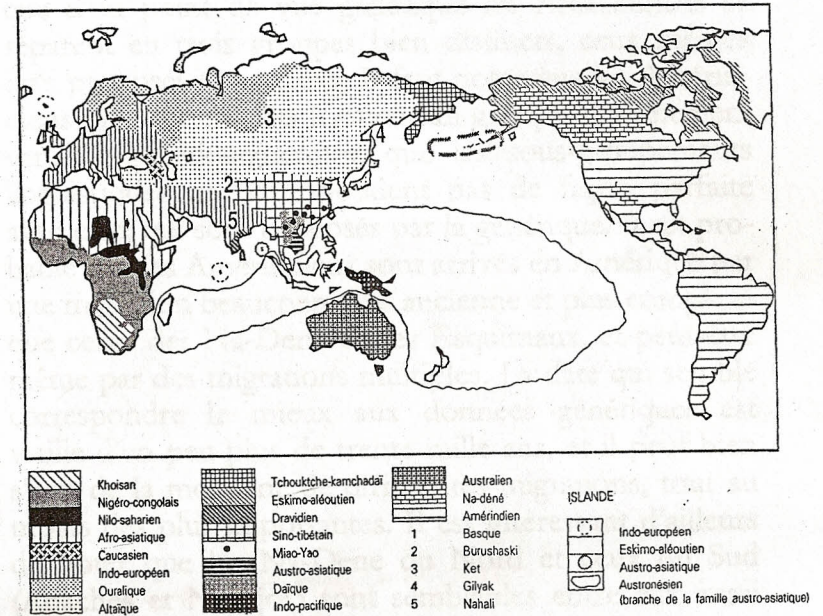


Figure 1. *Distribution géographique des familles linguistiques selon Ruhlen.*

Je montre dans la figure 1 la distribution géographique des familles suivant la classification proposée très récemment par M. Ruhlen, qui est à la fois la plus moderne et la plus complète. Pour donner quelque idée des différences d'opinion entre linguistes, disons qu'en général les familles de langues les moins connues, comme celles des Aborigènes d'Australie ou de Nouvelle-Guinée, sont considérées comme les plus difficiles et suscitent de nombreuses polémiques. Mais les opinions divergent aussi sur d'autres familles, et je voudrais citer un exemple de vraie discordance qui a donné lieu à une guerre presque centenaire, dont les cendres ne se

sont toujours pas éteintes, entre linguistes américains. Il s'agit des langues des Indiens d'Amérique.

Au début du siècle, un linguiste et un anthropologue célèbres, Sapir et Kroeber, ont tenté de montrer, sur la base de ressemblances communes à de nombreuses langues amérindiennes, que celles-ci pouvaient former un nombre restreint de familles. Ces affirmations ont rencontré des obstacles très puissants, car la plupart des linguistes américains se sont lancés dans une attaque violente contre cette unification. Un nouveau cycle a commencé sous l'impulsion de J. Greenberg, de l'université de Stanford, qui a publié un livre démontrant que les langues parlées par les Américains précolombiens peuvent se regrouper en trois familles : les langues esquimaudes, les langues na-dene, qui sont parlées surtout au nord de la côte pacifique, et les langues amérindiennes proprement dites, qui incluent la plupart des langues de ce continent. Greenberg a proposé également avec des biologistes américains que cette classification corresponde à celles qu'on peut obtenir sur les mêmes populations en utilisant soit les données des dents modernes et fossiles (C. Turner), soit celles des groupements sanguins et des protéines (Zegura). De plus, ces trois familles linguistiques semblent correspondre à trois migrations très importantes qu'on peut reconnaître archéologiquement et qui se sont succédé dans l'ordre inverse de l'ordre énuméré plus haut : les Amérindiens au début, les Na-Dene ensuite, et enfin les Esquimaux. Les premiers ont occupé toute l'Amérique, tandis que les deuxièmes et les troisièmes sont restés plus près de leur origine, la Sibérie. Nous avons confirmé les observations génétiques et nous trouvons

que d'un point de vue génétique les Amérindiens se séparent en trois groupes bien distincts, ceux mêmes qu'a proposés Greenberg. Il faut noter que les Amérindiens proprement dits forment un groupe extrêmement variable génétiquement et que les sous-groupements linguistiques ne correspondent pas de façon parfaite avec ceux qui sont proposés par la génétique. Il est probable que les Amérindiens sont arrivés en Amérique par une migration beaucoup plus ancienne et plus complexe que celles des Na-Dene et des Esquimaux, et peut-être même par des migrations multiples. La date qui semble correspondre le mieux aux données génétiques est vieille d'un peu plus de trente mille ans, et il peut bien s'agir de la moyenne de différentes migrations, tout au moins des plus importantes. Il est intéressant d'ailleurs de noter que les Na-Dene du Nord et ceux du Sud (Apaches et Navajos) sont semblables entre eux aussi du point de vue génétique, bien que les derniers montrent qu'ils ont absorbé une certaine quantité de flux génique des groupes amérindiens qu'ils ont traversés avant d'arriver à leur résidence actuelle.

La publication de ce livre a déchaîné presque immédiatement une nouvelle guerre des linguistes américains contre l'unité de la plupart des langues amérindiennes suggérée par Greenberg. Un groupe important de linguistes s'est réuni et a déclaré qu'il est impossible de reconnaître moins d'une soixantaine de groupements taxonomiques dans les langues amérindiennes. On sait bien que les taxonomistes sont souvent divisés en *lumpers* et en *splitters* : les tendances synthétiques et analytiques forment probablement une polarité de base dans l'esprit humain ; mais ici il y a des différences méthodo-

logiques que le livre de Greenberg discute en détail et qui peuvent expliquer une bonne partie de la dispute. Je ne suis pas linguiste, mais je peux suivre un discours général sur la méthode, et je trouve que les arguments de Greenberg sont convaincants. De plus, Greenberg a de son côté un avantage qu'on peut qualifier d'historique : il y a plus de trente ans, il a donné une nouvelle classification des langues africaines en quatre familles : la famille des langues afro-asiatiques, comprenant aussi toutes les langues sémitiques, celles qui sont parlées dans le nord de l'Afrique et la plupart des langues éthiopiennes ; la famille nilo-saharienne, qui comprend des langues parlées dans le haut Nil et le sud du Sahara ; les langues niger-kordofaniennes, qui incluent celles de l'Afrique de l'Ouest, du centre-ouest et du sud-est africain, y compris les langues bantoues qui ont eu récemment une grande diffusion ; et enfin les langues khoïsanides, parlées par les Hottentottes et les Bochimanes. La classification de Greenberg a soulevé au commencement une bagarre entre linguistes, mais à la fin elle a été acceptée sans réserves.

Il me semble intéressant de considérer quelques-unes des objections qui ont été faites à Greenberg par ses collègues américains, car elles aident à comprendre non seulement les difficultés objectives qui existent dans l'analyse de l'évolution linguistique, mais aussi celles, beaucoup moins objectives, qui ont été engendrées par les positions des adversaires de Greenberg. Les difficultés objectives tout d'abord : les langues évoluent très rapidement et il semble n'exister aucune ressemblance entre ces langues et d'autres langues plus éloignées. Le passage du temps est responsable de

changements phonologiques parfois très grands auxquels s'ajoutent souvent des changements sémantiques. La combinaison des deux phénomènes rend la reconstruction de ressemblances très difficile. La grammaire évolue aussi, bien qu'en moyenne plus lentement, et elle permet parfois de reconnaître plus aisément des ressemblances anciennes. Sous la pression de changements rapides, une langue devient incompréhensible dans un temps assez bref. Les langues issues du latin, parlées aujourd'hui en Europe, n'auraient pas été comprises des Romains il y a un millier d'années. Après une séparation de cinq ou dix mille ans, les mots qui montrent une parenté reconnaissable peuvent descendre à 5 ou 10 %. Heureusement, certains mots et certaines parties du discours montrent une vitesse d'évolution plus faible et les chances de reconstruire des parentés, même éloignées, sont alors plus grandes. Mais tout changement linguistique a des effets à d'autres niveaux, qui viennent compliquer l'analyse ; comme le disent les linguistes, dans une langue « tout se tient ».

Quant aux difficultés engendrées par des méthodologies erronées, certains linguistes du groupe américain anti-Greenberg ont soutenu qu'il était impossible de se déclarer pour une parenté sans une parfaite connaissance des langues examinées. Comme cela n'est possible que pour très peu de langues, on ne devrait jamais pouvoir se prononcer sur la parenté de plus de deux ou trois langues au maximum. En refusant d'utiliser la possibilité de mesurer la ressemblance entre deux langues et en réduisant le verdict à « apparenté/non apparenté », les linguistes américains du groupe anti-Greenberg ont du même coup tué la possibilité d'établir des hiérarchies classificatoires.

Il est intéressant que cette position soit totalement opposée à celle de plusieurs autres linguistes, qui mesurent la ressemblance entre langues sur la base du pourcentage de mots d'une liste standard ayant une origine commune reconnaissable. Cette méthode a été poussée plus loin par ses inventeurs, sur la base de l'hypothèse que la probabilité qu'un mot perde sa signification est constante dans le temps. En évaluant la fréquence de mots apparentés dans une période connue, par exemple dans la transition du latin aux langues modernes, on peut, grâce à cette hypothèse, calculer le temps nécessaire pour le développement de la différence existant entre deux langues modernes, et par conséquent évaluer la date approximative de leur séparation. Cette méthode, qui a pris le nom de « glottochronologie », fait donc appel à une « horloge linguistique » très proche intellectuellement de l'« horloge moléculaire ». En biologie, on a l'avantage de pouvoir utiliser beaucoup de protéines et d'acides nucléiques différents pour obtenir des évaluations indépendantes de la séparation entre deux espèces. On peut donc comparer des sources différentes d'information, ce qui permet un renforcement mutuel des conclusions. Malheureusement, c'est impossible en linguistique. La méthode est moins rigoureuse qu'en biologie et plus difficile à appliquer, surtout pour des périodes plus longues, puisque à partir de cinq à six mille ans de séparation entre deux langues le pourcentage de mots en commun est trop près de zéro pour être statistiquement fiable. La liste de mots ne peut pas être augmentée à loisir, car elle est déjà choisie parmi les mots qui changent le moins vite.

D'autres linguistes exigent que les ressemblances entre mots semblables de langues différentes soient examinées à la lumière des « correspondances classiques des sons », qui sont des règles très strictes de changements phonologiques observées d'une langue à l'autre. Selon ces linguistes, si les règles ne sont pas suivies exactement, il ne faut pas accepter la parenté de deux mots qui semblent à première vue avoir une origine commune. Or ces règles ont des exceptions, dont quelques-unes ont été expliquées par des sous-règles qui tiennent compte de l'« environnement » phonologique du changement. Mais il serait difficile, sinon impossible, d'expliquer toutes les exceptions car c'est en grande partie une question de probabilités de voies différentes de changement. Les « puristes » des correspondances classiques de sons n'acceptent pas les exceptions. Greenberg leur a donc répondu en établissant une liste impressionnante d'exceptions aux correspondances des sons qu'on trouve en comparant les langues de la famille indo-européenne. Il conclut qu'avec les règles que ses adversaires voudraient imposer, il eût été impossible d'établir la famille des langues indo-européennes, dont l'existence ne fait aucun doute, même s'il peut y avoir quelque discussion de détail. Fort heureusement, la famille indo-européenne a été établie avant la théorie de la correspondance des sons.

Enfin, des linguistes affirment qu'avant de démontrer des parentés entre groupements de langues, familles, sous-familles ou autres, il faudrait avoir reconstruit la « proto-langue », c'est-à-dire les mots du langage commun originel, qui est en général hypothétique. Ici on trouve une autre analogie avec la biolo-

gie, où l'on compare les séquences d'ADN en espèces différentes et où l'on cherche le « consensus », la séquence partagée par toutes les espèces comme probable séquence ancestrale. Mais la recherche du consensus est moins rigoureuse en linguistique car la variation y est beaucoup plus grande qu'en biologie et le consensus, après une évolution de quelques milliers d'années, n'est jamais susceptible de donner, avec une probabilité élevée, une représentation correcte du mot ancestral. En biologie, on peut trouver dans certaines protéines si importantes pour l'organisme que celui-ci ne pourrait tolérer de changement, exactement la même séquence d'acides aminés (les unités élémentaires dont les protéines sont faites) pourtant séparés par des milliards d'années. Il est vrai que la reconstruction des proto-langues peut faciliter l'analyse comparative, mais en imposant la nécessité de ce passage on limite sérieusement la portée de la méthode, puisque les proto-langues sont connues pour un petit nombre de familles seulement. De plus, certaines reconstructions sont probablement arbitraires ou ont une faible probabilité d'être correctes. La méthode de Greenberg évite ce passage. Elle a peut-être l'inconvénient de rester quelquefois plus subjective, mais elle peut aller beaucoup plus loin que les méthodes utilisées par les spécialistes des langues amérindiennes.

Dans les limites que nous nous sommes imposées ici, la classification de Ruhlen, un élève de Greenberg, me semble fournir des garanties satisfaisantes. La définition de « famille » ne paraît pas très sûre et il est clair qu'elle est subjective. Il n'est donc pas très important qu'une famille soit une sous-famille pour un auteur et

une super-famille pour un autre. Malheureusement, la classification moderne s'arrête au niveau des familles et celle de Ruhlen, qui en compte moins que les autres, en donne dix-sept. Il existe bien quelques superfamilles, mais les seuls moyens linguistiques ne suffisent pas à reconstituer un arbre complet partant d'une langue originelle commune à tous les hommes.

L'une de ces superfamilles présente un intérêt particulier. Il s'agit d'un groupement dont il existe aujourd'hui deux variantes : nostratique et eurasiatique. Son histoire est donnée par Ruhlen. La superfamille nostratique, comme elle a été décrite originellement par le linguiste russe Illyich-Svitic, puis par des scientifiques russes comme Dolgopolvsky, Gamkrelidze et Ivanov, et par d'autres encore, comprend les familles indo-européenne, ouralo-altaïque, afro-asiatique, dravidienne, caucasique. Shevoroshkin lui a rattaché la famille amérindienne définie par Greenberg. La superfamille eurasiatique définie par Greenberg est semblable à la nostratique, et s'en sépare en partie à cause des différentes définitions de quelques-unes des familles comme l'altaïque. Elle s'étend sur un espace un peu réduit par rapport à la nostratique, puisqu'elle n'inclut pas les familles afro-asiatique et dravidienne qui, selon Greenberg, ont des origines plus anciennes.

Avant de laisser de côté le sujet de la classification linguistique, je voudrais insister sur le fait que cinq ou six langues n'ont pas trouvé place dans les dix-sept familles de Ruhlen ou dans quelque autre classification. Un de ces « isolats » linguistiques nous est familier, c'est le basque. Parlée par environ vingt-cinq mille Français et un million et demi d'Espagnols, cette langue est vrai-

semblablement le reliquat d'une période pré-néolithique et descend peut-être des langues parlées par les Cro-Magnons, il y a vingt-cinq mille ans. Mais elle a certainement changé, de sorte que les Basques modernes et les hommes de Cro-Magnon, s'ils avaient la chance de dialoguer, ne se comprendraient pas et ne s'apercevraient probablement pas que leurs langues sont apparentées. Beaucoup de linguistes reconnaissent au basque une parenté avec des langues modernes du Caucase. Il est donc possible qu'une ou plusieurs familles de ces langues pré-indo-européennes fût parlée à l'époque paléolithique dans toute l'Europe, et peut-être même plus loin. D'autres linguistes voient des ressemblances entre les langues basques et caucasiennes, et les langues sino-tibétaines et na-dene de l'Amérique du Nord. Sur des bases que je ne peux juger, Ivanov affirme que les langues sumériennes, étrusques et d'autres langues vivantes éteintes sont également de la même famille, une famille ancienne qui s'étendait donc à tout le nord de l'Asie et de l'Europe.

L'inclusion du basque dans ce groupe fait penser que cette langue commune était parlée au moment où l'homme moderne est entré en Europe, il y a quarante mille ans. C'est un laps de temps très long pour l'évolution des langues. Le temps où le nostratique aurait commencé est évalué de dix mille à vingt mille ans. Quant à l'origine de la langue de l'homme moderne, on ne peut pas établir de dates sûres. Même pour l'origine, il y a une structure commune à toutes les familles, et cela est peut-être une considération suffisante pour penser que les langues parlées aujourd'hui ont assez d'origines communes pour prendre la date de disper-

sion de l'homme moderne à partir de l'Afrique comme point de départ des langues modernes. Cette date peut varier entre soixante-dix mille ans et cent cinquante mille ans. La première valeur est un minimum fixé par les dates d'hommes modernes les plus anciennes connues hors d'Afrique : en Chine, la première date est de soixante-sept mille ans, en accord avec la date attribuée à la première arrivée en Australie, il y a cinquante-cinq à soixante mille ans. Il est vrai que les premiers sites d'homme moderne au Moyen-Orient sont plus anciens, mais le fait de trouver l'homme de Neandertal dans le même espace à des dates postérieures a suggéré que la première colonisation moderne hors d'Afrique n'avait pas eu de succès. Les données biologiques sur la première séparation donnent appui à la limite supérieure : cent quarante-cinq mille ans est la date la plus récente (en février 1995) dérivée de l'ADN mitochondrial, et cent cinquante-cinq mille ans est celle à laquelle on arrive par l'ADN nucléaire.

Il n'existe pas d'arbre fiable de l'origine de toutes les langues, et il n'est même pas sûr qu'il y ait eu une origine unique du langage. La plupart des linguistes considèrent que ce problème est insoluble. C'est un peu comme de chercher à savoir s'il y a eu une seule origine de la vie sur terre. De nombreux biologistes pensent qu'il y a eu une origine unique, puisqu'il existe un seul type d'acides aminés synthétisés par les cellules vivantes, alors qu'il pourrait en exister deux, chimiquement opposés. Greenberg a noté qu'il y a au moins une étymologie commune à toutes les familles linguistiques, la racine *tik*, qui peut indiquer soit un doigt, soit le numéro un (une variation sémantique qui n'a pas besoin

d'être expliquée si on pense à la façon dont nous indiquons le numéro un avec un geste). Dans d'autres langues, on trouve d'autres changements sémantiques qui semblent encore acceptables, par exemple « main » ou même « bras ». En français, « doigt » vient de la racine latine *digit*; en italien, *dito*.

Sur la base de cet exemple, deux linguistes américains, Bengston et Ruhlen, ont proposé d'autres racines presque universelles. Mais il faudra sans doute beaucoup de temps avant que ces recherches très récentes trouvent l'accord des autres linguistes. Malheureusement, le nombre de mots sur lesquels on peut compter pour ces recherches est limité. Il s'agit en général de termes désignant les parties du corps, de pronoms personnels, de quelques autres noms et des chiffres un, deux, trois. Il n'est pas surprenant que les noms conservés depuis l'origine dans l'évolution linguistique soient ceux qu'on apprend les premiers : yeux, nez, bouche, etc. Mais il en existe d'autres qui étaient certainement très importants dans la vie paléolithique et qui le sont restés jusqu'aux temps les plus modernes, comme « pou ». Seule l'invention des insecticides modernes nous a libérés de ce fléau, et encore de façon incomplète.

Comparaison des familles linguistiques avec l'arbre génétique

Même s'il n'existe pas encore d'arbre évolutif complet des familles linguistiques, nous pouvons déjà comparer l'arbre génétique et ce qui existe de l'arbre lin-

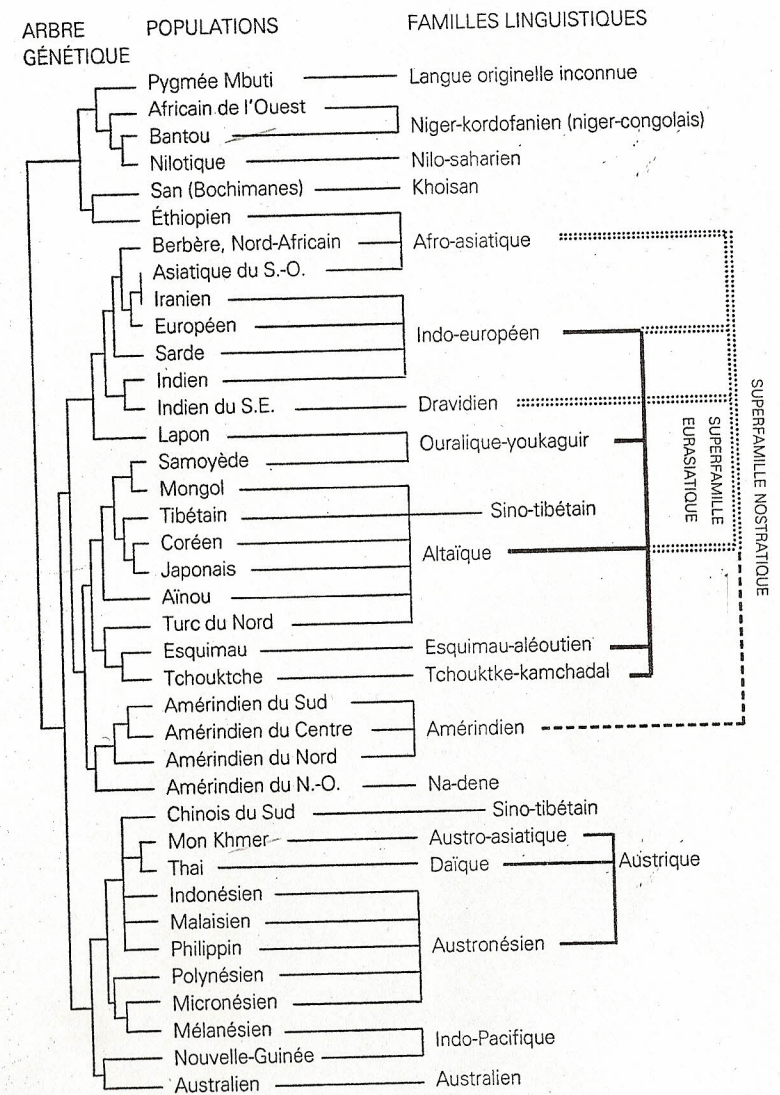


Figure 2. Arbre génétique et familles linguistiques.

guistique. En dépit de l'inachèvement de l'arbre des langues, on constate des ressemblances impressionnantes.

Dans la figure 2, les familles linguistiques ont été ajoutées près des populations qui parlent les langues appartenant à ces familles. On voit qu'une famille correspond souvent à une, deux ou trois branches de l'arbre génétique, rarement plus, et cela est vrai des seize familles linguistiques. La dix-septième famille, celle des langues du Caucase, n'est pas représentée. Il existe en effet deux familles de langues caucasiennes, septentrionale et méridionale, alors que Ruhlen, dans la première édition de son livre, n'en fait apparaître qu'une seule; mais il a accepté tout récemment cette subdivision.

Parfois, une famille linguistique correspond à une seule branche, car les populations parlant ces langues ont été groupées ensemble dans l'analyse génétique; elles montrent en effet une grande similarité, soit génétique, soit ethnographique, et sont géographiquement proches. Il s'agit par exemple de la sous-famille bantoue, une branche de la famille niger-kordofanienne, génétiquement très homogène par rapport aux autres Africains. Bien que le mot « bantou » dérive d'une classification linguistique, il est aussi utile comme terme de classification biologique. D'autres groupements linguistiques ont également été reconnus sur la base de données linguistiques, tels les Indiens du Sud qui parlent des langues dravidiennes et les Indiens américains qui parlent des langues na-dene. Ces populations utilisant des langues de la même famille sont très semblables, non seulement du point de vue linguistique, mais aussi du point de vue génétique et ethnique.

Comme l'arbre génétique montré dans la figure 2 comporte quarante-deux populations, parmi lesquelles certaines sont considérées comme « européennes » et groupées ensemble, et seize familles linguistiques, plusieurs populations de l'arbre génétique doivent se réunir dans une seule famille linguistique. Lorsque cela arrive, on constate que ce sont souvent des populations très proches dans l'arbre génétique qui parlent des langues de la même famille. On peut résumer ce comportement en disant que les familles linguistiques ont tendance à s'attacher dans la partie la plus basse de l'arbre, puisque les populations biologiques correspondantes ont une parenté génétique élevée. On pourrait donc utiliser l'arbre génétique pour dater approximativement l'origine d'une famille linguistique. On trouve alors que ces dates, à quelques exceptions près, sont comprises entre six mille et vingt-cinq mille ans selon les familles. On dispose de très peu d'informations linguistiques directes à ce sujet. La datation glottochronologique de la famille indo-européenne serait de six mille ans, mais une analyse statistique de Kruskal et ses collègues a montré que les estimations glottochronologiques de cet ordre de grandeur peuvent être sérieusement en dessous de la réalité.

Il y a toutefois des exceptions à la règle qui veut que les familles linguistiques comprennent toujours des populations très proches dans l'arbre génétique. Les Éthiopiens, par exemple, font partie de la branche génétique africaine, quoiqu'ils parlent des langues de la famille afro-asiatique, très répandue en Afrique du Nord, en Arabie et au Moyen-Orient où les hommes sont généralement caucasoïdes. Les Éthiopiens sont en

effet un peu plus africains que caucasoïdes sur le plan génétique et un peu plus caucasoïdes qu'africains sur le plan linguistique. Une autre exception de ce genre est celle des Lapons, qui font partie, génétiquement parlant, des caucasoïdes, dont ils se détachent par une branche très haute dans l'arbre de la figure 2. Ils parlent une langue de la famille ouralienne, dont on trouve des représentants typiques dans la Russie du Nord-Est et dans la Sibérie du Nord-Ouest, très près de l'Oural, d'où vient le nom de la famille. Les peuples ouraliens d'Asie sont génétiquement des Mongoloïdes, et les Lapons sont un mélange caucasoïde (probablement d'origine scandinave) et mongoloïde (d'origine sibérienne) avec prévalence du premier composant. Même sans regarder leurs gènes, qui confirment ce diagnostic, on peut le constater dans le changement de couleur de la peau, des cheveux et des yeux, qui varient du type mongoloïde au type scandinave.

Que ce soit dans le cas des Lapons ou des Éthiopiens, nous avons une explication simple de ce fait. Comme nous l'avons expliqué, ces deux populations viennent en effet de deux mélanges génétiques assez récents : les européens-sibériens pour les premiers, les africains-arabes pour les seconds. Les deux populations font partie des branches de l'arbre dont vient la majorité de leurs gènes, caucasoïdes pour les premiers et africains pour les autres, mais le mélange place ces populations à une position élevée de l'arbre. Il faut ajouter que si le mélange génétique montre exactement les proportions des deux populations ancestrales, on ne peut pas s'attendre à ce que les langues se mélangent entre elles avec la même simplicité que les gènes. Il est

clair que les peuples mêlés génétiquement tendent à garder seulement une des deux langues correspondantes aux populations d'origine. Parfois, la langue de la population mêlée ne montre aucune altération ; plus souvent, on trouve des mots ou quelquefois des sons empruntés à d'autres langues. La grammaire résiste au mélange encore mieux que le dictionnaire. Quant à l'histoire des deux mélanges, nous savons que des contacts très intimes ont eu lieu entre Arabes et Africains en Éthiopie au cours du premier millénaire avant et après Jésus-Christ : un empire arabo-éthiopien a eu sa capitale d'abord à Saba, en Arabie, puis à Aksoum, en Éthiopie ; mais probablement les contacts avaient commencé avant cette période. Nous savons aussi que les Lapons sont restés dans la région qu'ils habitent aujourd'hui pendant au moins deux mille ans. Dans les deux cas, un laps de temps suffisant s'est écoulé pour que des échanges génétiques aient lieu entre les populations en contact. Les évaluations du mélange génétique varient selon la population considérée, en Laponie et en Éthiopie. Mais il est facile de calculer que le génome originel d'un peuple qui reçoit un flux génique constant de 5 % par génération de ses voisins gardera seulement 70 % de son génome après trois siècles. C'est exactement le cas des Noirs américains, qui ont été importés comme esclaves en Amérique dès la fin du XVI^e siècle et ont connu un flux génique blanc de cet ordre de grandeur pendant ces trois siècles. Si ce flux continuait à la même vitesse, les Noirs américains ne garderaient pas plus de 10 % de leurs gènes originels mille ans après leur arrivée en Amérique. Les Lapons et les Éthiopiens, qui ont été en contact longtemps (plusieurs milliers d'années),

ont gardé une plus grande partie de leurs gènes originaires puisque évidemment ils ont un flux génique plus faible que celui de leurs voisins.

On trouve encore d'autres exceptions à la correspondance exacte entre les arbres génétique et linguistique, qui sont intéressantes car elles « confirment la règle ». Ainsi, les Tibétains se trouvent génétiquement dans le groupe des Mongoloïdes du Nord ; ils parlent des langues sino-tibétaines, comme les Chinois, mais les Chinois, dans notre arbre, viennent du sud de la Chine et sont génétiquement plus semblables aux Mongoloïdes du Sud. Dans ce cas, l'histoire va venir à notre secours. Selon les historiens, les Tibétains sont chinois d'origine nord-chinoise. Partis au III^e siècle avant Jésus-Christ de la Chine du Nord, où ils étaient des pasteurs nomades, ils se sont portés au sud, au Tibet. Plusieurs d'entre eux sont encore des pasteurs nomades. Les Tibétains ont gardé leurs langues originelles après leur migration vers le sud. L'unification de la Chine eut lieu vers le III^e siècle avant Jésus-Christ. Commencée par la dynastie Qing, assez brève, elle fut achevée par la dynastie Han, qui régna pendant quatre siècles successifs. Les deux dynasties régnaient au nord et furent responsables de l'extension de leur langue, originaire de la Chine du Nord, à presque tout l'empire chinois. Dans les deux mille ans qui suivirent, cette langue naturellement se différença en plusieurs autres langues. En Chine, surtout au sud, on trouve quand même un grand nombre de minorités ethniques (dont cinquante-cinq sont reconnues officiellement), et chacune d'entre elles a conservé sa langue originelle, issue de nombreuses origines différentes.

Les tentatives de reconstruction d'une partie de l'arbre linguistique faites par la création des superfamilles nostratique et eurasiatique, avec l'addition de la famille des Indiens d'Amérique, sont représentées graphiquement à l'extrême droite de l'ensemble des familles linguistiques, dans l'arbre de figure 2. Il n'y a pas d'accord complet entre les deux superfamilles. Elles ont été reconstruites par des chercheurs indépendants qui utilisaient un matériel et une méthode différents, mais elles représentent une unité de base et il est très probable que les recherches futures pourront expliquer ces différences. Si on considère l'ensemble des deux familles nostratique et eurasiatique, leur union correspond à peu d'exceptions près à la grande branche génétique que nous avons appelée nord-eurasienne et qui réunit Caucasoïdes, Mongoloïdes du Nord et Indiens d'Amérique. Cette branche se forme lors de la deuxième scission, lorsque les non-Africains se séparent entre la branche Europe-nord de l'Asie et la branche Sud-Est-asiatique (comprenant aussi l'Australie et la Nouvelle-Guinée). Les exceptions les plus importantes à l'identité de la branche Europe-nord de l'Asie et de l'union nostratique-eurasiatique sont basques, caucasiques, sino-tibétaines, etc. Elles sont comprises dans la branche génétique, mais non dans la branche linguistique. C'est exactement la très ancienne famille eurasiatique, qui aurait vécu au temps de l'expansion en Europe de l'homme moderne et qui a été presque complètement oblitérée par la superposition plus tardive de l'eurasiatique-nostratique. Une analyse plus détaillée des données génétiques montre que ce sont surtout les Na-Dene méridionaux, les Apaches et les

Navajos, qui se sont fortement mêlés aux Indiens du nord de l'Amérique. Ce phénomène est en partie responsable de la tendance des Na-Dene à s'associer dans l'arbre avec les autres Indiens d'Amérique, de nouveau en position haute.

L'origine du groupe linguistique Na-Dene n'est pas bien connue. Quelques linguistes mettent la famille en relation avec la famille sino-tibétaine, dont l'origine nous est également inconnue, mais cette dernière a dû se répandre du nord au sud de la Chine. Les Chinois du Nord, qui ne sont pas représentés dans l'arbre à cause de l'insuffisance de données génétiques, sont très semblables génétiquement aux Mongols, aux Coréens et aux Japonais ; autrement dit, ils sont des Mongoloïdes du Nord. La présence d'une langue sino-tibétaine dans la branche génétique du Sud-Est asiatique (les Chinois du Sud) est la conséquence de l'obligation faite il y a deux mille ans aux Chinois par les empereurs de la dynastie Han de parler la langue de la Chine du Nord. Cette dynastie donna son nom au mot *Han*, par lequel la majorité chinoise se distingue des ethnies minoritaires.

Il faut aussi ajouter que l'arbre génétique montré dans la figure 2 a des points faibles. Le premier et le plus robuste des embranchements représente la division entre les Africains et les autres, conséquence de la première diaspora, et se retrouve avec n'importe quelle méthode permettant de calculer les distances génétiques ou de reconstruire l'arbre. Dans notre arbre de 1988, la deuxième branche sépare l'Océanie et l'Asie du Sud-Est des autres continents. En 1994, il devint clair que d'autres distances donnaient la préférence à la sépara-

tion entre l'Océanie (Nouvelle-Guinée et Aborigènes d'Australie) et le reste, qui était subdivisé entre Asie du Sud-Est et le reste de l'Eurasie dans un embranchement successif. Les mélanges qui eurent lieu en Asie sont probablement la cause des difficultés non encore complètement résolues à propos de ces branches. L'arbre que nous montrons ici est un peu différent de ceux que nous avons montrés auparavant. Il faudra encore beaucoup de données supplémentaires sur davantage de gènes et de populations pour arriver à une version définitive, surtout en ce qui concerne les rapports entre les trois branches majeures qui se forment après l'embranchement Africains-non-Africains.

Des difficultés ultérieures sont intervenues parce qu'une méthode de reconstruction d'arbres très populaire aujourd'hui, développée par Saitou et Nei et dite *neighbour joining*, a donné un arbre dans lequel les branches qui succèdent à la première sont très différentes de la nôtre. Après l'embranchement Africains-non-Africains, qui reste le premier, la première branche est l'Europe, mais elle est très courte. L'explication la plus plausible est que les Européens ont reçu de l'Asie leur principale contribution génétique ; mais à cause de leur proximité avec l'Afrique, ils ont eu aussi des échanges avec ce continent (dont on trouve les traces les plus nettes en Afrique du Nord). Dans la représentation du *neighbour joining*, le mélange des populations cause un fort raccourcissement des branches et leur déplacement vers l'origine de l'arbre. Je crois que la différence entre les arbres obtenus avec cette méthode et la méthode plus traditionnelle que nous avons employée tire son origine de ce fait et que le *neighbour*

joining est très sensible aux mélanges, en présence desquels il produit des modifications grossières par rapport à l'histoire.

Le mélange entre l'Asie du Nord et l'Afrique contribue à la genèse de l'Europe, en particulier pour la première. Plusieurs caractères génétiques des Européens font qu'ils sont à peu près intermédiaires entre les Africains et les Asiatiques de l'Est. Le Moyen-Orient est un bon candidat pour l'aire d'origine des Européens. Intermédiaire entre l'Asie et l'Afrique, il a probablement été le point de départ de l'occupation de l'Europe par l'homme moderne, il y a quarante mille ans, et par les hommes néolithiques, il y a dix mille ans. On ne peut pas exclure que, pour occuper l'Europe, l'homme moderne soit venu en partie de plus haut, de la région des steppes. Ce sont probablement ces mouvements qui ont porté en Eurasie la famille linguistique (dite eurasienne pour la distinguer de l'eurasiatique, qui est apparue plus tard) dont les seules familles couvrant encore un espace important aujourd'hui sont la sino-tibétaine et la na-dene. Les autres familles sont réduites à des isolats : le basque en Europe (et peut-être une langue éteinte, l'étrusque), les langues du Caucase du Nord, un isolat dans les montagnes du Pakistan (le bouroushaski, parlé par les Hunza). La famille eurasienne a été remplacée dans le Nord par la famille eurasiatique, qui a donné naissance aux cinq familles des temps modernes. Dans le Sud, elle a été remplacée par l'afro-asiatique et la dravidienne, qui se sont diffusées dans le nord de l'Afrique, dans le sud-ouest de l'Asie et au Moyen-Orient. Aujourd'hui, la dravidienne, qui était autrefois parlée en Perse, au Pakistan et en Inde, est limitée au

sud de la péninsule indienne et à quelques isolats du nord. Elle a été remplacée en Perse, au Pakistan et dans la plus grande partie de l'Inde par les langues indo-européennes.

Pourquoi est-il légitime de s'attendre à une similarité de base entre l'évolution biologique et l'évolution linguistique ?

Des analogies importantes existent entre l'évolution des gènes et celle des langues. Dans les deux cas, un changement qui se manifeste d'abord chez un individu peut ensuite se répandre dans toute la population. Dans le cas des gènes, on appelle ce changement *mutation* ; il passe successivement par la génération d'un individu à un autre et peut mettre un certain temps, parfois des millions de générations, pour qu'on le retrouve chez tous les individus d'une population. C'est que le génome, le patrimoine héréditaire est bien caché et protégé des influences extérieures ; il change rarement et la transmission d'un individu à l'autre se fait seulement du père ou de la mère à l'enfant. Les changements du langage, eux, sont bien plus fréquents et peuvent passer d'un individu à l'autre sans qu'il y ait aucun lien de parenté entre ces individus. Il est facile de comprendre que les langues peuvent changer beaucoup plus vite que les génomes. En effet, si un mot peut durer mille ans, un gène peut rester inchangé pendant des millions et parfois même des milliards d'années. Malgré ces différences, il reste des raisons de similarité importante entre les deux évolutions.

Mais il faut tout de suite préciser qu'il n'y a aucune raison de penser que les gènes influencent la possibilité de parler l'une ou l'autre langue. L'homme moderne possède au départ la capacité d'apprendre n'importe quelle langue connue, et la langue maternelle est le résultat d'un accident individuel : le lieu et groupe social de naissance. Toutes les langues modernes ont une complexité de structure comparable, et les langues des groupes ethniques qui vivent à un niveau économique primitif ne sont pas du tout plus « primitives » que les nôtres. S'il y a un effet d'interaction entre gènes et langues, ce sont plutôt les langues qui influencent les gènes dans la mesure où une différence de langue entre populations peut diminuer les échanges génétiques sans toutefois les annuler.

L'évolution linguistique est un cas particulier d'évolution culturelle et nous lui consacrerons le prochain chapitre. Ce qui nous intéresse ici, c'est de comprendre comment il est possible de trouver un parallélisme entre deux évolutions si différentes l'une de l'autre. L'explication est fort simple. Deux populations isolées l'une de l'autre se différencient soit du point de vue génétique, soit du point de vue linguistique. L'isolement, qu'il soit dû à des barrières géographiques, écologiques ou sociales, empêche et rend moins probables les mariages entre les deux populations, et par conséquent l'échange génétique. Les populations vont donc évoluer indépendamment et, de ce fait, se différencier l'une de l'autre. La différenciation génétique va augmenter régulièrement avec le passage du temps. On peut attendre exactement la même chose d'un point de vue linguistique : l'isolement va diminuer ou annuler

les échanges culturels et les deux langues vont aussi se différencier. Même si les résultats de l'estimation du temps de séparation de deux langues par la glottochronologie ne sont pas toujours aussi exacts qu'on le voudrait, ce temps augmente généralement avec la durée de l'isolement. En principe, l'arbre linguistique et l'arbre génétique doivent correspondre puisqu'ils reflètent la même histoire de scissions et donc d'isolements évolutifs.

Il existe toutefois des causes importantes de perturbations. Lors d'événements historiques particuliers, un langage peut être remplacé par un autre. En Europe, par exemple, la langue hongroise se trouve géographiquement au centre d'un groupe de langues indo-européennes de différentes branches : slave, germanique, romaine ; mais elle appartient à la famille ouralienne, branche ugro-finnoise. Les autres langues de la même famille sont parlées dans le nord-est de l'Europe et dans l'ouest de la Sibérie, comme nous l'avons vu en parlant des Lapons. C'est à la fin du IX^e siècle après Jésus-Christ que les Magyars, un groupe de nomades, ont laissé leurs demeures en Russie pour traverser les Carpates et envahir la Hongrie, qui avait déjà été occupée en partie par d'autres barbares, les Avars. La conquête a abouti à la création d'une monarchie magyare qui a imposé la langue des vainqueurs à la population de langue romaine vivant dans la région. L'effectif des conquérants était plus faible que celui des conquis, et représentait à peu près 30 % du total. L'effet génétique de la conquête a donc été modeste et dilué par les échanges successifs avec les voisins. Aujourd'hui, un peu plus de 10 % des gènes des Hongrois

peuvent être attribués aux conquérants d'origine ouralienne.

Ailleurs, les conquêtes des barbares après la chute de l'Empire romain ont eu plus de difficultés à remplacer ou même seulement à altérer de façon importante la langue des vaincus, qui étaient toujours supérieurs en effectif et en degré de civilisation bien que leur organisation militaire ait été incapable de résister à la poussée des barbares. Les Lombards, venus probablement de Suède, ont occupé une bonne partie de l'Italie un peu avant l'invasion de la Hongrie et ont constitué un état puissant qui a duré jusqu'au XI^e siècle sans avoir toutefois d'effets importants sur la langue. Cela est vrai aussi des Francs, une population d'origine allemande qui a eu une très grande importance dans l'histoire politique de la France sans en affecter la langue. Mais en Angleterre, après la chute de l'Empire romain, des Anglo-Saxons, mercenaires des Romains, ont probablement fait venir du continent leurs familles et leurs alliés, et ont pu imposer leur langue vers le VI^e siècle avant Jésus-Christ. Les îles Britanniques ont connu en peu de temps de grands changements de langues. Les indigènes parlaient des langues aujourd'hui inconnues ; au cours du dernier millénaire avant Jésus-Christ, la langue celtique s'est répandue dans presque toute l'Europe, à partir d'une zone d'origine située probablement entre l'Autriche et la Suisse. Au temps de la conquête romaine, les îles Britanniques parlaient des dialectes celtiques. La conquête des Romains leur a substitué la langue latine, suivie par la langue anglo-saxonne, et enfin la conquête normande a importé dans l'anglais moderne un grand nombre de mots d'origine française.

Un autre remplacement important s'est produit en Turquie, où les Turcs ont commencé à attaquer l'empire de Byzance à la fin du XI^e siècle et ont terminé leur conquête avec la chute de Constantinople en 1453. Le remplacement de la langue grecque, qui était parlée à Byzance, par la langue turque, était particulièrement important puisque cette langue appartient à une famille différente, dite altaïque. De nouveau, l'effet génétique a été modeste en Turquie. Les armées de cette époque étaient composées d'un nombre assez restreint de soldats et, même si parfois ils voyageaient avec leur famille, les effectifs étaient toujours faibles par rapport à ceux des pays envahis lorsque ceux-ci avaient une longue histoire de civilisation et de développement économique. Mais protégés par l'Empire romain, les peuples envahis avaient perdu la capacité de résister à des envahisseurs dangereux comme les barbares.

Nous avons évoqué à plusieurs reprises une autre langue parlée en Europe qui n'est pas indo-européenne mais est parlée aujourd'hui encore par une proportion de Français du Sud-Ouest et une plus grande partie de la population de l'Espagne du Nord : la langue locale, le basque, n'a pas été remplacée par une langue indo-européenne, arrivée plus tard, à la différence de ce qui s'est passé ailleurs.

Les exemples de remplacement de langues ne sont pas limités à l'Europe ; mais comme l'histoire de l'Europe est assez longue, les derniers remplacements se sont produits à une période historique, tandis qu'ailleurs il s'est agi d'événements presque toujours préhistoriques. Les invasions ariennes de l'Iran, du Pakistan et de l'Inde ont apporté des langues indo-européennes

dans des régions où l'on parlait des langues de famille dravidiennes. Du sud-est de l'Asie, les grandes découvertes géographiques des malayo-polynésiens ont apporté les langues austronésiennes en Nouvelle-Guinée, en Mélanésie, Micronésie et Polynésie. Du côté occidental, ces langues sont parvenues jusqu'à Madagascar, où on les parle encore. Les migrations polynésiennes ont eu moins d'effet génétique là où il y avait des densités élevées de Mélanésiens (en Mélanésie); dans cet endroit, il y a eu plus de trois mille ans de mélanges et la mosaïque génétique-linguistique est très complexe. Mais lorsqu'il y a trois mille ans les Austronésiens sont passés de la Polynésie centrale à la Polynésie orientale, ils devaient être encore des Mongoloïdes, car ils n'avaient pas eu beaucoup de temps pour se mêler aux Mélanésiens.

Les amateurs d'explorations extraordinaires seront contents de savoir qu'on ne peut pas exclure à partir d'une base génétique que l'Amérique du Sud ait donné une contribution partielle à la Polynésie orientale, comme cela a été suggéré par Thor Heyerdahl avec les voyages du *Kon-Tiki*. La différence génétique entre Mongoloïdes et Amérindiens est si faible qu'il est encore assez difficile de dire précisément quelle a été la contribution génétique des Sud-Américains à la Polynésie. Les nouveaux marqueurs génétiques qui viennent d'être introduits pourront sans doute donner des réponses beaucoup plus claires à tous ces problèmes.

La substitution totale d'une langue se fait plus facilement sous une forte pression politique organisée. Ces circonstances extraordinaires font que maintes fois les interactions entre les peuples ne produisent pas de

changements de langues. Les langues différentes parlées dans les pays voisins peuvent être peu affectées pendant des milliers d'années, même si les gènes subissent une substitution partielle qui va parfois jusqu'à une substitution totale. Il est difficile de préciser l'importance de la substitution du génome basque originel à celui des voisins dû au flux génique, mais elle doit être assez grande. Dans d'autres cas, la substitution génique est presque complète, tandis que la substitution linguistique ne l'est pas. Deux populations parlant des langues khoisanides en Tanzanie (Hadza, Sandawe) n'ont plus de gènes des Khoisans; mais les Hadzas et les Sandawes ne sont pas très nombreux et doivent avoir vécu complètement entourés de Bantous pendant plus de mille ans. Cette période de faible échange génétique est suffisante pour atteindre un niveau de remplacement de plus de 90 % des gènes originaux. Les Hadzas et les Sandawes étaient des chasseurs-récolteurs et ont donc été séparés des agriculteurs bantous par des divergences socio-économiques de manière suffisante pour avoir conservé leur langue, mais insuffisante pour bannir tout échange génétique avec les voisins.

Pour résumer, le remplacement des langues n'est pas la seule cause de perturbation du parallélisme entre évolutions génétique et linguistique observé aujourd'hui. Les gènes aussi peuvent être remplacés. Les voisins géographiques favorisent les échanges génétiques, même si ces échanges ont surtout lieu entre des voisins très proches, là où les groupes sociaux et politiques se touchent et se mêlent.

Il est remarquable que malgré les chances de remplacement génétique et linguistique on puisse encore

trouver dans la mosaïque linguistique et génétique du monde moderne assez de cohérence pour pouvoir reconstruire un arbre commun aux deux évolutions.

*Les grandes expansions humaines
et les familles linguistiques*

Nous l'avons dit, l'origine de la plupart des familles linguistiques, à en juger par l'arbre génétique, semble dater de six mille à vingt-cinq mille ans. Quelques familles sont plus anciennes : celle des langues australiennes et des langues indo-pacifiques (Nouvelle-Guinée) peuvent dater de plus de quarante mille ans sur la base de la première entrée de l'homme moderne dans ces parties du monde. Ici, la définition de la famille est aidée par l'isolement géographique des deux grandes îles, Australie et Nouvelle-Guinée. Les deux familles correspondantes sont différentes l'une de l'autre, de même que les Aborigènes le sont d'un point de vue génétique entre eux. Les langues khoisanides doivent être anciennes, mais il est plus difficile de leur donner un âge. La « famille eurasiennne », postulée sur la base des ressemblances entre basque, langues caucasiennes, sino-tibétaines, na-dene, burushaski, et autres langues isolées ou éteintes, était peut-être au maximum de son extension au temps de l'occupation de l'Europe par l'homme moderne, il y a quarante mille ans.

La famille la mieux connue reste l'indo-européenne. On a essayé de lui trouver un lieu d'origine, avec des résultats incroyablement différents. D'un auteur à l'autre, on a proposé de nombreux espaces, de

l'Allemagne jusqu'au nord-est du Caucase, des États baltiques jusqu'à Suez. C'est un espace aussi large que celui couvert par la famille et dans certains cas il la déborde même. Il y a peu de temps encore, l'une des théories les plus populaires était celle de l'archéologue Marjia Gimbutas, qui place leur origine au-dessus de la mer Noire et associe les premiers peuples parlant des langues européennes avec la civilisation des « kourganes » dans les steppes asiatiques. Mais lorsque Gimbutas a publié son hypothèse, les dates des kourganes étaient mal connues. Elle a parlé de trois mille à trois mille cinq cents ans avant Jésus-Christ, une date qui passe pour impossible, car trop ancienne, aux yeux des archéologues anglais. Il semble que ces dates aient été confirmées très récemment par de nouvelles fouilles qui ont aussi montré que les chevaux étaient probablement domestiqués et montés, et que les chars de guerre étaient bâtis dans cette région.

L'archéologue Colin Renfrew a avancé en 1987 une autre hypothèse, très intéressante, selon laquelle les langues indo-européennes auraient été diffusées par les agriculteurs néolithiques à partir du Moyen-Orient. J'ai déjà eu l'occasion de mentionner ce livre, dans le chapitre 4, puisqu'il a corroboré notre hypothèse, à savoir que la diffusion de l'agriculture néolithique a été un phénomène démique et non culturel.

Il était difficile de ne pas noter la correspondance entre la diffusion des langues indo-européennes et une partie de l'aire de diffusion de l'agriculture ; mais dans mes discussions avec Ammermann, mon collaborateur archéologue dans ce travail, nous avons décidé que le travail archéologique ne pouvait pas nous informer sur

la langue lorsque l'écriture elle-même était absente. Un archéologue du niveau de Renfrew peut se permettre d'ignorer la règle. Lorsque je lui avais rendu visite, il y a quelques années, il m'avait donné les bases de la théorie sur laquelle il était en train d'écrire son livre. Je lui avais fait remarquer la difficulté qui existait dans les aires méridionales de l'expansion de l'agriculture du Moyen-Orient, où l'on parle des langues de la famille afro-asiatique qui ont une origine différente. Après la parution de son livre, en 1987, Renfrew a ajusté son tir en tenant compte des nouvelles connaissances linguistiques à propos de la famille des langues dravidiennes, qui ont fait qu'elle a été appelée elamo-dravidienne. Il a été possible de montrer que, dans la région d'Elam, dont la capitale est Soussa, au sud-ouest de l'Iran, la langue parlée et écrite en cunéiforme appartenait à la famille dravidienne. Ce fait, dont la publication du livre de Ruhlen (1987) a permis de diffuser la connaissance, donne aussi d'autres informations intéressantes à ce sujet : il existe aujourd'hui encore des langues dravidiennes au Pakistan et dans l'Inde du Nord, autres que celles qui sont parlées dans toute la partie méridionale de la péninsule indienne. La diffusion des langues dravidiennes devait être jadis beaucoup plus vaste que celle montrée par leur distribution moderne, limitée au sud de la péninsule indienne.

Ce fait est par ailleurs l'objet d'une coïncidence remarquable : j'ai publié dans la revue *Muniba*, en 1988, à l'occasion d'une conférence sur les Basques, une figure qui résumait une hypothèse selon laquelle, au moment du développement de l'agriculture, on parlait au Moyen-Orient trois langues différentes qui se sont

diffusées dans trois directions avec les agriculteurs et ont donné naissance à trois familles linguistiques. Or, à peu près au même moment, Renfrew publiait une correction de son hypothèse de 1987 dans la revue *Phil. Transactions*, dont je vis le résultat pour la première fois dans un article qu'il publia en 1990 dans la revue *Scientific American*. Les figures que nous avons indépendamment dessinées sont extrêmement semblables. La question qui nous intéresse ici est la proposition par Renfrew que les langues indo-européennes ont eu leur origine en Turquie, puis se sont répandues en Europe avec les paysans néolithiques. Il est clair que tous les immigrants emportent leur langue avec eux, et s'ils ne rencontrent aucun habitant dans les terres qu'ils vont occuper, ils n'ont pas de raison de changer leur langue. Il vaut la peine d'ajouter que les peuples d'Europe, avant l'arrivée des hommes néolithiques (souvent appelés mésolithiques), étaient dans la plupart des régions en faible densité, et comme ils vivaient de chasse et de récolte, ils choisissaient de préférence des environnements de forêt, différents géologiquement de ceux qui étaient meilleurs pour la culture.

L'idée de Renfrew donne aussi une date à la radiation des langues indo-européennes : sept mille à sept mille cinq cents ans avant Jésus-Christ, soit l'époque du départ des agriculteurs. Cette date est une des causes de la résistance que l'idée a rencontrée, car des évaluations glottochronologiques, quoique très approximatives, auraient donné un âge de six mille ans, ce qui est en accord avec l'origine kourgane suggérée par Gimbutas. Je suis convaincu, quand même, que les idées de Gimbutas et de Renfrew sont justes l'une et l'autre. Il n'y a

pas de contradiction entre elles ; au contraire, elles se renforcent mutuellement. La date glottochronologique sur l'origine de l'indo-européen n'est pas fiable, puisque la relation exponentielle sur laquelle elle se base n'est pas valable sur tout l'intervalle (Kruskal, 1971). D'autres objections d'ordre écologique, comme le nom des arbres de la région d'origine, qui semblent plus favorables à une origine kourgane, sont fiables seulement pour cette région ; ce qui explique probablement la plupart des langues européennes.

Il est clair que, génétiquement parlant, les peuples de la steppe sont sûrement des descendants des néolithiques du Moyen-Orient qui ont immigré dans la région des steppes en se dirigeant vers le nord, passant à l'ouest ou à l'est de la mer Noire. Après leur arrivée, ils ont domestiqué le cheval qu'on ne trouvait pas partout avec la même abondance et cela leur a permis de survivre et de prospérer dans un environnement hostile à l'agriculture. L'adaptation a pris du temps mais avec le développement des armes en bronze, qui a commencé à cette époque, vers cinq mille ou six mille ans avant Jésus-Christ, ils se sont trouvés prêts pour une expansion : ils avaient de la nourriture, des moyens de transport, des armes nouvelles et puissantes. La région est devenue pendant trois ou quatre mille ans le point de départ de nombreuses expansions qui se dirigèrent vers le sud-est, par l'Asie centrale, en direction de la Perse, l'Afghanistan, le Pakistan et l'Inde ; et vers l'ouest, en direction de l'Europe centrale et de l'Europe du Nord.

On le voit, ces deux expansions successives ont eu des origines et des directions différentes, mais elles ont toujours concerné les mêmes peuples. Les langues des

kourganes avaient entre trois mille cinq cents et quatre mille ans de différence de celles qui étaient originaires du Moyen-Orient, et étaient donc déjà différenciées les unes des autres. Les langues indo-européennes, arrivées des steppes eurasiatiques avec la culture arienne, ont remplacé presque totalement les langues dravidiennes en Iran, au Pakistan et dans l'Inde septentrionale, mais à la périphérie de l'expansion arienne les langues indo-européennes ne sont pas arrivées à s'imposer tout à fait : les langues dravidiennes sont encore parlées de nos jours, en particulier dans le sud de la péninsule indienne.

Il s'avère que les gènes suivent l'évolution générale des langues. Tous les peuples de l'Inde sont caucasoïdes, même s'ils sont plus sombres de peau que les Caucasoïdes du Nord, mais les populations du Sud qui parlent des langues dravidiennes ont une composition génique un peu différente. Dans cette partie du monde, au moins trois couches ethniques se superposent : la plus ancienne et la plus limitée, qui n'est malheureusement pas bien étudiée, montre un type physique ayant quelque ressemblance avec les aborigènes australiens (les Australoïdes) ; puis viennent les Dravidiens, surtout au sud de l'Inde où ils ont été moins mêlés aux envahisseurs ariens ; les derniers, les Ariens, sont plus clairs de peau. Dans la partie centrale et orientale de l'Inde, on a eu aussi des infiltrations du sud-est de l'Asie. Une analyse des composantes principales des cartes géniques de l'Inde, non encore publiée mais établie en collaboration avec Madhav Gadgil, de Bangalore, ses collaborateurs et Eric Minch, de Stanford, montre que la première composante atteint un pic vers

l'Afghanistan, où les Ariens sont passés, et montre une forte corrélation avec les dates archéologiques qui prouve la première arrivée du cheval dans les différentes régions indiennes. Il n'existait pas de chevaux sauvages en Inde, cet animal y ayant été introduit par les Ariens.

L'association des deux idées de Gimbutas et de Renfrew me semble beaucoup plus vraisemblable que celle d'une origine de toutes les langues indo-européennes modernes qui proviendrait seulement de la Turquie, comme le suggère la deuxième hypothèse. Cette idée nous vient d'une nouvelle étude. Les langues indo-européennes ont été analysées très soigneusement dès le siècle passé après la naissance officielle de la famille, qu'on peut dater de 1786. Cette année-là, le juge anglais Sir William Jones lançait dans une fameuse conférence tenue à la Bengal Asiatic Society de Calcutta, qu'il venait de fonder et dont il était le président, la théorie que les langues sanscrite, grecque et latine, et peut-être aussi les langues gothique et celtique, montraient des ressemblances telles qu'il fallait admettre qu'elles avaient eu une origine commune. Auparavant, d'autres savants avaient déjà noté des ressemblances entre le sanscrit et les langues européennes, tels le marchand florentin Filippo Sassetti au *xvi*^e siècle et le père jésuite Coerdoux qui, à la fin du *xviii*^e siècle, avait envoyé de Pondichéry à l'Académie des Inscriptions de Paris des notes montrant que sanscrit, grec et latin devaient avoir une origine commune. Mais ces observations n'ont pas eu le retentissement de la conférence de Jones. En 1863, le linguiste allemand August Schleicher a publié un arbre de l'origine des langues indo-

européennes qui est assez semblable à celui qu'on peut dessiner aujourd'hui en utilisant des méthodes modernes. Mais à cette époque il n'y avait pas de méthodes quantitatives – elles ont été développées beaucoup plus tard, dans la seconde moitié du *xx*^e siècle – pour des applications biologiques. Toutefois, le rapport entre biologie et linguistique a été assez étroit, même au début. Schleicher a certainement été influencé par l'usage théorique des arbres fait par Charles Darwin pour expliquer la théorie de descente des organismes, et Darwin, dans le chapitre *xiv* de *L'Origine des espèces*, publié en 1855, dit clairement que si on connaissait l'arbre de descente biologique des groupes humains on pourrait en tirer celui des langues. Cet essai, comme nous l'avons vu, n'a été fait qu'en 1988 pour la première fois et indépendamment de la prophétie de Darwin.

La première analyse quantitative moderne et complète de la ressemblance entre les langues indo-européennes ne date que de 1992. Elle est due au statisticien Kruskal et aux linguistes Dyen et Black, qui ont mesuré la fréquence des mots d'origine commune dans toutes les paires possibles entre environ six douzaines de langues indo-européennes, en utilisant la liste glottochronologique standard de deux cents mots. En appliquant à leurs données deux méthodes modernes de reconstruction d'arbres développées en génétique, nous avons obtenu des arbres reproductibles qui correspondent approximativement à l'arbre de Schleicher. La plus grande différence est la position de la racine. Les groupements plus importants de langues indo-européennes sont les sous-familles germanique (qui

comprend aussi la langue anglaise et les langues scandinaves), italique (issue du latin), balto-slave, celtique, indienne et iranienne (que la plupart des linguistes considèrent comme un seul branchement, tandis que, selon Kruskal et ses collègues, elles sont séparées). Plusieurs langues ont une origine isolée et séparée de ces sous-familles : l'albanais, l'arménien et, plus tard, le grec. Les langues éteintes comme l'hittite et le tocharien ne sont pas comprises dans cette analyse. Le même arbre a été obtenu par deux méthodes de reconstruction (UPGMA et NJ) et est donné en résumé dans la figure 3.

Il est possible que les langues isolées comme l'albanais et l'arménien et, avec moins d'évidence, le grec soient originaires de la première vague, qui est partie de la Turquie. Leur ancienneté par rapport aux autres branches est responsable de leur position dans l'arbre. Elles sont aussi géographiquement plus proches de la Turquie.

Notre analyse rassemble les langues indienne et iranienne dans une branche indo-iranienne, selon la tradition classique des études indo-européennes et à la différence des conclusions de Kruskal et al., obtenues par une méthode statistique qui n'utilise pas les arbres. Mais la partie commune aux branches, dans notre arbre, est brève et sa longueur n'est pas significativement différente de zéro, ce qui diminue la force de la conclusion selon laquelle les langues indiennes et iraniennes ont une origine commune.

Les autres branches sont peut-être dérivées de la deuxième vague de migrations indo-européennes, qui part de l'aire kourgane de l'ouest, tandis que la branche indo-iranienne viendrait d'une aire kourgane de l'est.

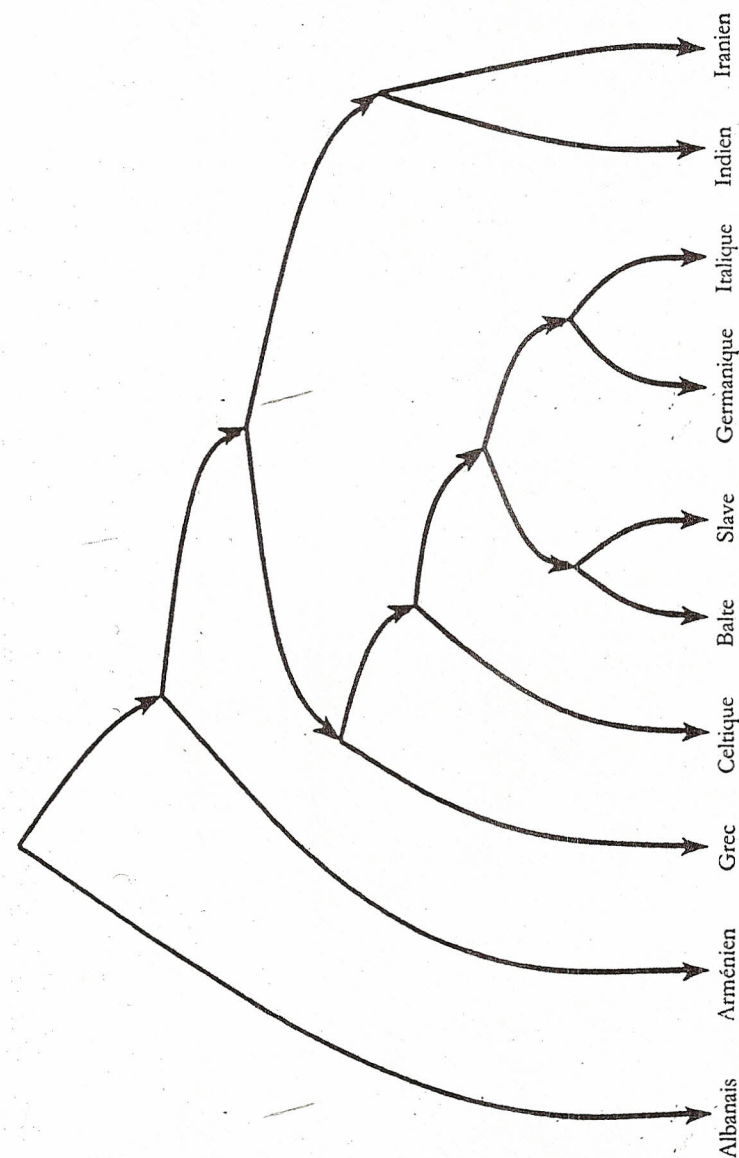


Figure 3. Résumé de l'arbre des langues indo-européennes.

Bien d'autres expansions ont apporté des langues nouvelles à des terres nouvelles. Les expansions démiques que nous connaissons ont presque toutes été accompagnées par leur langue d'origine. Entre les expansions préhistoriques étudiées soit génétiquement, soit linguistiquement, celle des populations parlant la langue bantoue présente un intérêt particulier. Malgré les contacts et les échanges avec des peuplades parlant d'autres langues, comme les Nilo-Sahariens en Afrique de l'Est et les Khoisanides en Afrique du Sud, les Bantous ont gardé leur individualité génétique, qui les distingue des autres Africains de l'Ouest dont ils sont issus. Partis du Nigeria et du Cameroun, ils se sont dirigés vers le sud, non loin de la côte atlantique. Cette première expansion a débuté à l'époque néolithique, il y a plus de trois mille ans. Plus tard, l'expansion bantoue a été favorisée par l'utilisation du fer, dont on trouve des traces archéologiques très claires au Nigeria dès cinq cents ans avant Jésus-Christ. C'est seulement vers l'an zéro de notre ère que les Bantous ont atteint la région des Grands Lacs (Ouganda et Kenya) et, dès cette époque, l'archéologie trouve un fort développement de l'usage du fer.

L'expansion bantoue vers le sud s'est développée du côté occidental et du côté oriental, les deux courants se mêlant plus tard. Il semble que les Bantous étaient encore à quelques centaines de kilomètres du Cap lorsque les Hollandais y ont débarqué vers 1650. La vitesse de cette expansion est un peu plus rapide (environ un kilomètre et demi par an) que celle de l'agriculture néolithique qui est partie du Moyen-Orient. Dans la plupart de leurs expansions, les Bantous pou-

vaient utiliser des haches de fer, plus efficaces que les haches de pierre pour ouvrir la forêt. La génétique montre que le sud de l'Afrique a été atteint d'abord par l'expansion occidentale et que l'expansion orientale s'est arrêtée bien avant, ce que prouvent à la fois les dernières découvertes archéologiques, qui montrent que les Bantous sont arrivés très tôt dans la région de la Namibie, et les plus récents résultats de la linguistique, qui n'avait pas jusqu'ici exploré suffisamment les langues de la branche occidentale.

Il n'est pas possible que la corrélation entre langues et gènes soit parfaite puisque les conquêtes rapides de grandes régions peuvent permettre de remplacer les langues originelles par des langues non apparentées, mais ces phénomènes ne semblent pas assez fréquents pour que la corrélation disparaisse. Nous voyons également qu'en cas d'échanges génétiques prolongés avec des voisins différents, les gènes peuvent être remplacés. Cependant, malgré ces deux causes de confusion, la corrélation entre gènes et langues reste positive et statistiquement significative.

Il est intéressant de noter que même au niveau microgéographique, les régions ayant été soumises à une étude détaillée ont montré des corrélations très élevées entre géographie, génétique, linguistique et autres aspects culturels, comme les noms de famille. Souvent, la mosaïque génético-linguistique observée nous montre très clairement les effets des nombreuses expansions – certaines sont connues historiquement –, leurs superpositions et leurs interactions. Il arrive que l'on trouve des perturbations, mais elles ne parviennent pas dans la plupart des cas à détruire la clarté de la corrélation entre peuples, gènes et langues.

L'évolution des langues est un sujet du plus haut intérêt. Dans ce chapitre, nous n'avons fait que l'aborder pour expliquer les raisons des ressemblances entre gènes et langues. Mais l'évolution linguistique est aussi très importante pour comprendre l'évolution culturelle, que nous allons analyser dans le chapitre suivant.

Chapitre 6

LA TRANSMISSION ET L'ÉVOLUTION CULTURELLES

L'homme se distingue des autres animaux, même de ses cousins les plus proches, par la richesse de sa culture et par la place qu'elle occupe dans sa vie. La culture n'est pas réservée à la seule espèce humaine comme on ne manque pas de le constater dès que l'on accepte de définir le terme de façon assez large. Les anthropologues ont proposé des centaines de définitions de la culture qui sont pour la plupart très abstraites et excluent la technologie. Je préfère adopter une position inverse et donner la définition la plus simple et la plus vaste possible : elle inclut la technologie, qui a joué et continue de jouer un rôle essentiel dans l'évolution culturelle et dans l'histoire humaine. Une telle définition inclut la culture animale, naturellement moins développée que la culture humaine puisque la communication est très réduite, mais dont les origines et les conséquences sont identiques.

La culture est le résultat de ce qu'on apprend d'autrui, par opposition à ce qu'on apprend seul, par soi-même. S'il s'agit dans les deux cas d'un apprentissage, la différence est grande entre apprendre par soi-