



HAL
open science

Origine des langues et du langage

Bernard Victorri

► **To cite this version:**

Bernard Victorri. Origine des langues et du langage. Encyclopædia Universalis, 2006. halshs-00009947

HAL Id: halshs-00009947

<https://shs.hal.science/halshs-00009947>

Submitted on 4 Apr 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Origine des langues et du langage

Bernard Victorri

On dissocie généralement la question de l'origine des langues et du langage en deux problématiques distinctes. L'étude de l'origine des langues consiste à établir une généalogie des langues humaines en identifiant leurs relations de parenté à partir des langues actuelles et de langues anciennes ayant laissé des traces écrites. L'objectif est de remonter le plus loin possible dans le temps, la question ultime étant de savoir si toutes les langues humaines dérivent d'une seule et même langue originelle, ou si elles proviennent au contraire de langues ancestrales ayant émergé indépendamment les unes des autres. Le questionnement sur l'origine du langage est tout autre : il s'agit de comprendre comment notre espèce a acquis la faculté de langage en étudiant l'évolution des hominidés. A quel moment de cette évolution le langage est-il apparu ? Y a-t-il eu des étapes intermédiaires et de quelle nature ? Quelles pressions évolutives ont conduit à l'émergence de cette forme de communication ?

Les deux problématiques diffèrent donc tant par l'objectif des recherches que par les méthodes utilisées et les disciplines convoquées. Même la question de l'unicité de l'origine se pose différemment. En effet, on peut très bien défendre la thèse d'une polygenèse du langage en même temps que celle d'une langue mère unique : le langage aurait émergé à plusieurs reprises dans l'histoire de l'homínisation, mais toutes les langues connues seraient issues d'une même langue ancestrale, toutes les langues issues d'autres sources ayant disparu sans laisser de descendance.

Néanmoins, on assiste ces dernières années à un rapprochement entre ces deux grandes problématiques. En effet, au fur et à mesure que progresse notre compréhension des dernières étapes de l'homínisation et de l'histoire des premières migrations humaines, notamment grâce à l'apport de la génétique des populations, se mettent en place des scénarios qui pourraient rendre compte à la fois de l'origine des langues et du langage. Ces scénarios sont encore très spéculatifs, mais les travaux pluridisciplinaires en cours, qui mobilisent des chercheurs d'horizons les plus divers, depuis l'éthologie animale et la neurobiologie jusqu'à la modélisation informatique et les sciences cognitives, en passant par l'archéologie et la paléanthropologie, font espérer de nouveaux progrès qui viendront resserrer encore davantage l'éventail des scénarios plausibles, en confirmant ou infirmant certaines des hypothèses qui les sous-tendent.

La démarche de la linguistique historique

La méthode essentielle de la linguistique historique est le comparatisme : on recherche dans les langues dont on veut établir la parenté des similarités grammaticales, comme par exemple des analogies dans les systèmes de conjugaisons ou de déclinaisons, et des similarités lexicales, en établissant des listes des mots qui se correspondent de langue à langue, tant au plan phonétique que sémantique. Si ces correspondances sont suffisamment systématiques, on en déduit que les langues font partie d'une même famille qui résulte de l'évolution différenciée d'une même langue ancestrale. Les mots qui se correspondent dans ces langues sont alors appelés des cognats.

Le premier grand succès de la méthode comparatiste a été la découverte au 19^{ème} siècle de la famille indo-européenne. On a en effet établi la parenté d'un vaste ensemble de langues, comprenant des langues indiennes comme le sanscrit et l'hindi, les langues iraniennes (persan, kurde, ossète, etc.), les langues romanes (le latin et ses descendants français, espagnol, italien, roumain, etc.), les langues germaniques (dont l'anglais, l'allemand, le danois), les langues slaves (dont le russe, le polonais, le serbo-croate), les langues celtiques (irlandais, gallois, gaulois, breton, etc.), et des langues plus esseulées, notamment le grec, l'arménien, et l'albanais.

La méthode permet aussi, quand les données sont suffisantes, de reconstruire la langue ancestrale, que l'on appelle la protolangue associée à la famille. Le procédé consiste à découvrir les lois d'évolution phonétique qui ont présidé à la différenciation des langues, et qui sont donc à l'origine des lois de correspondance phonétique mises en évidence lors de la comparaison des langues de la famille. Par exemple, la comparaison entre les langues romanes et les langues germaniques fait apparaître une correspondance systématique entre la consonne /p/ en roman et la consonne /f/ en germanique, comme on peut le constater sur les couples suivants de cognats italiens et anglais : *padre* – *father* (« père »), *pesce* – *fish* (« poisson »), *piede* – *foot* (« pied »), etc. Mais cette loi de correspondance phonétique ne nous dit pas quelle était la consonne des mots du proto-indo-européen à l'origine de ces cognats. Pour le découvrir, il faut analyser l'ensemble des langues de la famille. Ainsi pour « père », on observe que dans la plupart des branches de l'indo-européen c'est un /p/ qui apparaît à l'initiale : *pitar* en sanscrit, *pater* en grec, *pater* en latin, etc. On en déduit donc que /p/ devait être la consonne originelle, et que le /f/ que l'on observe dans les langues germaniques est le résultat d'une loi d'évolution qui a transformé /p/ en /f/ dans cette branche

particulière de la famille. On peut alors reconstruire la forme des mots correspondants en proto-indo-européen.

La méthode comparatiste a permis, au cours du 20^{ème} siècle, d'identifier partout dans le monde des familles de langues du type de la famille indo-européenne. A chacune de ces familles correspond en principe une protolange, dont descendent toutes les langues actuelles de la famille. Certaines de ces familles comprennent plusieurs centaines de langues, comme la famille bantoue, dont les langues se répartissent sur une bonne moitié du continent africain. D'autres sont beaucoup plus restreintes, le cas extrême étant celui de certaines langues qui restent inclassables, comme le basque par exemple : on les appelle des isolats. On peut estimer à quelques centaines le nombre de ces familles, isolats compris. Pour un grand nombre de ces familles, un travail de reconstruction de la protolange a été entrepris, avec les mêmes méthodes que pour le proto-indo-européen. Même si les progrès sont très inégaux suivant les familles, on peut donc dire qu'une première étape a été accomplie dans l'établissement d'une généalogie des cinq à six mille langues parlées dans le monde.

Du point de vue temporel, ce travail correspond à une remontée dans le temps de quelques milliers d'années. C'est ainsi que l'on estime que le proto-bantou a dû être parlé il y a moins de deux mille ans. On doit noter cependant de très grandes disparités dans la profondeur temporelle atteinte suivant les familles. C'est sans aucun doute avec la famille indo-européenne que l'on a pu aller le plus loin, parce que l'on dispose d'un nombre important de données écrites provenant de langues anciennes comme le latin, le grec et le sanscrit (mais aussi le gotique, le hittite, etc.), qui datent elles-mêmes de plus de deux mille ans (et même plus de trois mille ans pour le sanscrit védique). On estime généralement que le proto-indo-européen a dû être parlé il y a quelque six mille ans, voire plus. A l'inverse, il est clair que cette profondeur est nulle pour les isolats, puisqu'il est impossible de reconstruire une protolange à partir de la donnée d'une seule langue attestée.

Les macrofamilles

Certains linguistes ont cherché à remonter plus avant dans le temps : ils ont tenté de regrouper des familles entre elles, créant ainsi des macrofamilles, en comparant les protolangues de ces familles, ou à défaut, en utilisant des caractères très répandus dans une famille quand la protolange correspondante n'a pas été reconstruite. Les données sont beaucoup plus ténues, et l'argumentation moins solide. Néanmoins, des hypothèses très intéressantes ont été avancées. Dès les années 1960, l'école linguistique russe a défendu la

thèse d'une macrofamille appelée le nostratique, qui regrouperait, pour l'essentiel, outre la famille indo-européenne, les familles ouraliennes (finnois, hongrois, estonien, same, etc. ainsi qu'un groupe de langues sibériennes), altaïque (comprenant notamment le turc, le mongol et le mandchou), eskimo-aléoute (dont les langues inuits), ainsi que la famille sémitique (arabe, hébreu, araméen, amharique, etc.). Parallèlement, Joseph Greenberg proposait de regrouper les familles de langues africaines en quatre macrofamilles : khoïsan (Afrique australe : langues bochimanes, entre autres), nilo-saharien (Afrique subsaharienne : massaï, songhaï, langues nubienne, ...), nigéro-kordofanien (la plus grande partie de l'Afrique, avec la grande famille bantoue, mais aussi le wolof, le peul, le bambara, etc.), et enfin afro-asiatique, comprenant notamment l'haoussa, les langues berbères, les langues couchitiques (oromo, somali, afar), les langues chamitiques (égyptien des hiéroglyphes, copte) ainsi que les langues sémitiques. Dans les années 1980 d'autres grands regroupements ont été avancés. Joseph Greenberg classe les langues américaines autochtones en trois grands ensembles : une macrofamille amérindienne, comprenant toutes les langues d'Amérique du Sud et un grand nombre d'Amérique du Nord (comme les langues algonquines, sioux et iroquoises), la famille na-déne (langues de l'Alaska et langues apaches comme le navajo), et la famille eskimo-aléoute. Comme on l'a vu, cette dernière famille avait été regroupée avec l'indo-européen et d'autres dans le nostratique. Greenberg reprend cette idée en proposant une macrofamille eurasiatique, qui n'a pas exactement la même composition que le nostratique (en particulier elle ne contient pas la famille sémitique, que Greenberg, on l'a vu, place plutôt dans la macrofamille afro-asiatique ; on y ajoute aussi des isolats comme le japonais, le coréen et l'aïnou). Quant à la famille na-déne, plusieurs auteurs la rapprochent des langues du Nord-Caucase (comme le tchéchène et l'abkhaze), qui elles-mêmes ont été associées aux langues sino-tibétaines, le tout formant donc une macrofamille appelée déne-caucasien, qui, selon certains, comprendrait aussi des isolats comme le basque, le bouroushaski et le sumérien. Sur la base de l'ensemble de ces travaux, Merritt Ruhlen, disciple de Greenberg, a tenté plus récemment une vaste synthèse dans laquelle il regroupe toutes les langues du monde en douze macrofamilles. Outre les sept déjà évoquées (les quatre africaines, l'amérindienne, l'eurasiatique et le déne-caucasien), on y trouve le kartvélien (langues du Sud-Caucase, dont le géorgien), le dravidien (langues de l'Inde du Sud et de l'Est, comme le tamoul et le télougou), l'austroasiatique (regroupant quatre familles : la famille austro-asiatique, dont le vietnamien et le khmer, la famille austronésienne avec notamment le malais, le javanais, le tagalog, le malgache et un millier de langues de Taiwan et d'Océanie, la famille tai-kadai composée de langues d'Asie du Sud-Est, dont le thaï, et la famille miao-yao, également présente en Asie du Sud-Est),

l'australien (quelques centaines de langues australiennes) et l'indo-pacifique (une dizaine de familles, dont les langues papoues, les langues des Iles Andaman ou encore celles des Iles Salomon).

Ce classement en douze macrofamilles est extrêmement controversé. Beaucoup de linguistes critiquent le principe même de ces regroupements. D'autres proposent des regroupements différents, avec des arguments parfois plus convaincants. On peut néanmoins espérer que la poursuite de ces travaux de comparaison entre familles finiront, au moins dans certains cas, à aboutir à des conclusions mieux étayées qui feront consensus. Si cet espoir n'est pas vain, cela voudra dire que la linguistique historique peut remonter dans le temps bien au-delà de la barrière des six mille ans qui a longtemps semblé infranchissable, puisqu'un « proto-nostratique » ou un « proto-eurasiatique » aurait été parlé plusieurs milliers d'années avant le proto-indo-européen qui en dériverait.

La langue originelle

Peut-on espérer aller encore plus loin, et démontrer que ces macrofamilles sont elles-mêmes issues d'une seule langue mère ? C'est cette dernière étape que n'a pas hésité à franchir Ruhlen, en collaboration avec John Bengtson. Ils ont cherché des ressemblances entre les macrofamilles, en se centrant sur le vocabulaire de base (des mots comme *eau*, *ciel*, *homme*, les parties du corps, les chiffres un et deux, etc., qui sont les plus résistants au changement linguistique). Comme on ne dispose pas de reconstitution des lexiques des protolangues correspondant à ces macrofamilles, ils ont pris pour base 32 familles (dont des isolats comme le basque), largement représentatives des douze macrofamilles. Ils ont ainsi établi une trentaine de racines « mondiales », dont chacune est présente dans une proportion importante des familles étudiées, et qui feraient donc partie du lexique de la langue mère originelle. Ruhlen écrit même : « A long terme, nous sommes sûrs que la masse de données témoignant de la monogenèse des langues existantes deviendra si contraignante que la question ne sera plus de savoir si toutes les langues du monde sont apparentées, mais pourquoi il a fallu si longtemps à la communauté linguistique pour s'en apercevoir ».

Mais cet enthousiasme n'a pas suffi à convaincre : la plupart des linguistes sont restés très sceptiques sur la valeur scientifique de ce travail. Quand on observe de près les données présentées, on constate effectivement que les similitudes relevées, tant au plan phonétique que sémantique, sont très fragiles, et pourraient, pour une bonne part, être simplement dues au hasard. Prenons l'exemple de la racine AQ'WA « eau », que Ruhlen affirme retrouver dans 13

des 32 familles. Au plan phonologique, on ne retrouve dans certaines familles qu'une consonne (k, g, q, ...) qui ressemble au Q' originel ; de même, en ce qui concerne le sens, c'est tout un champ sémantique qui est mis à contribution : à côté de « eau », on trouve « liquide », « humide », « goutte », « pluie », « nuage », « rivière », « lac », « boire », « sucer », « laver »... Chacun de ces changements, pris isolément, est plausible : ce sont des évolutions tout à fait classiques, tant pour les sons que pour les sens. Mais rien ne prouve que le mot *aqua* qui veut effectivement dire « eau » en latin soit vraiment un cognat du mot *k'ā* qui veut dire « boire » dans certaines langues khoïsanés, ni du mot **yoka* reconstruit du proto-ouralien avec le sens de « rivière », ni de *oko* qui veut dire « pluie » dans certaines langues amérindiennes ! On peut montrer au contraire que l'on a de bonnes chances de trouver dans n'importe quelle famille de langues un mot de ce vaste champ sémantique présentant une des consonnes adéquates, sans avoir à postuler une quelconque relation de parenté. En revanche, on remarquera que les mots latins *aqua* et français *eau* n'ont plus un seul phonème en commun alors que leur parenté est certaine et très récente !

Plus généralement, il ne semble pas rationnel d'espérer trouver des ressemblances phonologiques entre cognats de langues ou proto-langues ayant divergé il y a si longtemps. Rappelons que ce sont des lois de correspondance phonétique qu'il faut établir, et non des ressemblances, qui, elles, ont toute chance d'être trompeuses. Pour reprendre l'exemple de l'italien et de l'anglais, si l'on peut soupçonner que *piede* et *foot* sont des cognats, ce n'est pas parce qu'ils se ressemblent, mais parce que le /p/ de l'un correspond au /f/ de l'autre (ainsi d'ailleurs que le /d/ et le /t/). D'autres mots de ces deux langues se ressemblent bien plus, comme par exemple *pasta* et *paste* (« pâte »), ou encore *famiglia* et *family* (« famille »). En fait, c'est la preuve que leur parenté ne remonte pas au proto-indo-européen, puisque l'on a pas la correspondance /p/ - /f/. Effectivement, cette parenté est bien plus récente : ce sont des mots d'origine latine, qui ont enrichi le lexique du moyen anglais il y a moins d'un millénaire, par l'intermédiaire d'emprunts massifs au moyen français et au franco-normand, à la suite de la conquête de l'Angleterre par Guillaume le Conquérant.

Ainsi semble-t-il tout à fait illusoire, malgré les prétentions de Ruhlen, de demander à la linguistique historique de répondre à la question de la langue originelle. Et cela d'autant plus qu'une langue mère unique éventuelle remonterait à plus de cinquante mille ans. En effet la différenciation en macrofamilles a forcément débuté avec la dispersion géographique des humains à travers le globe à partir du berceau africain, et l'on sait notamment que la colonisation de l'Australie par notre espèce date d'au moins cinquante mille ans.

En revanche, la linguistique historique peut très bien contribuer à une meilleure connaissance de l'histoire des migrations de nos ancêtres à des dates plus récentes. Par exemple, parallèlement à la classification de Greenberg des langues américaines en trois grands groupes (macrofamille amérindienne, famille na-dene et famille eskimo-aléoute), des études de génétique des populations, menées notamment par Luca Cavalli-Sforza, ont montré que les distances génétiques entre les populations parlant ces langues étaient nettement plus grandes entre ces trois groupes qu'à l'intérieur de chacun des groupes. Les résultats concordants de ces deux disciplines permettent donc de conforter un des scénarios de l'histoire du peuplement de l'Amérique, celui qui postule trois grandes vagues de migration successives en provenance de la Sibérie : la première, correspondant à l'amérindienne, aurait abouti à la colonisation de l'ensemble du continent, jusqu'au fin fond de l'Amérique du Sud, tandis que la deuxième, correspondant au na-dene, n'aurait touché que l'Amérique du Nord, et que la troisième, la plus récente, aurait été cantonnée aux régions septentrionales proches de la Sibérie, où l'on parle effectivement les langues eskimo-aléoutes.

La linguistique historique a donc un rôle important à jouer dans cette grande entreprise pluridisciplinaire qui, en regroupant les efforts des linguistes, des paléo-anthropologues, des archéologues et des généticiens des populations, vise à reconstituer l'histoire de l'humanité depuis l'apparition de notre espèce.

En ce qui concerne la question de l'unicité de la langue originelle, elle sera peut-être élucidée autrement, à partir des travaux consacrés à l'origine du langage. En effet, un grand nombre de chercheurs défendent la thèse d'un « goulot d'étranglement » dans l'histoire de l'hominisation : tous les humains actuels descendraient d'un petit groupe d'*Homo sapiens* ayant vécu il y a cent à deux cent mille ans. Cette thèse ne fait pas l'unanimité, mais c'est elle qui, aujourd'hui, explique le mieux les données des différentes disciplines impliquées, notamment la génétique des populations et la paléogénétique (étude du matériel génétique fossile). Si ces chercheurs ne se trompent pas, l'existence d'une langue mère unique est alors très plausible : on peut supposer que ce petit groupe possédait déjà le langage humain, et donc, vraisemblablement, une seule et même langue, puisque, selon les estimations des généticiens, il devait compter au plus quelques milliers d'individus. Toutes les langues du monde proviendraient donc *in fine* de cette langue, qui se serait diversifiée au cours de la dispersion de notre espèce. L'autre possibilité, à savoir que le langage soit apparu dans notre espèce après la dispersion, donc en différents endroits de manière indépendante, semble peu crédible. En effet, même si l'on n'adhère pas à la thèse chomskyenne d'une « grammaire

universelle » innée, on admet le plus souvent que l'aptitude des humains à acquérir une première langue réclame des dispositions innées, qui sont les mêmes pour tous les humains et pour toutes les langues (un nouveau-né de n'importe quelle origine acquière sans difficulté la langue de la communauté dans laquelle il est élevé, quelle qu'elle soit). Ce fait semble peu compatible avec la thèse d'une polygenèse du langage au sein même de notre espèce, qui impliquerait des différences génétiques de la faculté d'acquisition du langage que l'on pourrait corrélérer avec certains types de langues.

Ainsi, la langue mère aurait bien existé, même si sa reconstitution par les méthodes de la linguistique historique semble impossible en l'état actuel de nos connaissances.

L'apparition du langage

Si la plupart des chercheurs pensent que la faculté de langage n'est pas apparue *après* l'émergence de notre espèce, le consensus est moins net sur la question de savoir si elle est apparue *avant*, autrement dit, si d'autres espèces d'hominidés, comme notamment les Néandertaliens, possédaient un langage de même nature que le nôtre.

Le processus d'hominisation, qui a duré six à huit millions d'années, a comporté de nombreuses étapes. Deux stades sont généralement considérés comme les plus pertinents dans les discussions sur l'émergence et l'évolution du langage : l'apparition d'*Homo erectus*, il y a plus d'un million d'années, et la naissance de notre propre espèce, *Homo sapiens* moderne, il y a quelque cent mille ans. *Homo erectus* a représenté un incontestable succès évolutif : premier hominidé à sortir d'Afrique, il a conquis tout l'Ancien Monde, et il a connu un développement continu de ses facultés cognitives et de sa technologie, conduisant ainsi à la naissance des différentes lignées d'*Homo sapiens* archaïques (comme l'homme de Neanderthal en Europe, l'homme de Rhodésie en Afrique, l'homme de Solo à Java, ou, découvert plus récemment aussi en Indonésie et encore sujet à controverse, l'homme de Flores), qui maîtrisaient le feu, la chasse aux gros mammifères, la construction d'habitats, etc. Quant à notre espèce, issue de l'un de ces groupes d'*Homo sapiens* archaïques en Afrique de l'Est, elle s'est à son tour dispersée sur toute la planète, supplantant les autres descendants d'*Homo erectus*, qui auraient fini par disparaître entièrement (il y a environ trente mille ans pour les Néandertaliens, plus récemment encore pour l'homme de Flores), sans se mêler à notre espèce, si l'on en croit la thèse du goulot d'étranglement. C'est à des représentants de notre espèce que l'on doit ce que l'on a appelé « l'explosion symbolique », qui a commencé il

Il y a quelque trente mille ans, et dont l'art pariétal (grottes ornées comme la grotte Chauvet ou la grotte de Lascaux par exemple) représente l'expression la plus spectaculaire.

La plupart des chercheurs qui travaillent aujourd'hui au problème de l'origine du langage s'accordent à penser que le système de communication des hominidés s'est lui aussi développé en deux étapes, qu'il est tentant de faire correspondre respectivement aux succès d'*Homo erectus* et d'*Homo sapiens* moderne, même si l'on ne dispose d'aucun élément vraiment probant qui pourrait valider complètement cette thèse.

On a cru dans les années 1980 pouvoir utiliser des données fossiles sur la conformation anatomique de l'appareil phonatoire. On sait en effet que les humains ont un conduit vocal très différent de celui des grands singes, avec en particulier à une position plus basse du larynx, qui permet de proférer une très grande variété de sons, mais qui présente aussi un inconvénient majeur : on peut s'étouffer en avalant « de travers » (obstruction de la trachée artère par de la nourriture). La descente du larynx n'a donc pu devenir un avantage évolutif qu'en raison d'une pression de sélection liée au langage. On a donc cherché à reconstruire la forme du conduit vocal de différents hominidés à partir d'éléments osseux fossiles (notamment la forme de l'os hyoïde). Les premières simulations, réalisées par Philip Lieberman, semblaient prouver que les premiers *Homo erectus* n'étaient pas très différents des grands singes de ce point de vue, et que l'homme de Neanderthal lui-même n'avait pas la capacité de produire toute l'étendue des sons du langage humain. Mais l'interprétation de ces derniers résultats a été infirmée par d'autres simulations. De plus, il n'est pas dit qu'un répertoire sonore aussi développé que le nôtre soit un prérequis pour l'émergence du langage humain, et, à vrai dire, l'existence des langues des signes, utilisées par les sourds, montre que le langage humain peut se passer de l'ensemble de l'appareil phonatoire !

Les données sur la taille et la conformation du cerveau ont aussi été mises en relation avec l'évolution du langage. Le coefficient d'encéphalisation (qui mesure le volume de l'encéphale relativement à la taille corporelle) a connu une forte augmentation au cours des derniers cinq cent mille ans chez *Homo erectus*, pour atteindre une taille comparable à la nôtre chez les différents *Homo sapiens* archaïques (voire plus grande chez les Néandertaliens). Même si, là encore, rien ne prouve que le degré de complexité du langage soit directement corrélé à la taille du cerveau, l'ensemble de ces données conduit à penser que l'on peut faire effectivement remonter la première étape du développement du langage à plusieurs centaines de milliers d'années, au cours de l'évolution d'*Homo erectus*, la deuxième étape ayant eu lieu beaucoup plus récemment, chez les *Homo sapiens* archaïques, sans que l'on puisse dire si elle

n'a concerné que notre propre espèce, ou si d'autres espèces d'*Homo sapiens* archaïques, aujourd'hui disparues, ont connu une semblable évolution.

Le protolangage

C'est le linguiste Dereck Bickerton qui a proposé le premier d'appeler protolangage le système de communication qui aurait précédé le langage humain, dans le cadre de l'hypothèse d'une évolution du langage en deux étapes. Il faut souligner que ce protolangage n'a rien à voir avec les protolangues dont nous avons parlé ci-dessus : contrairement à celles-ci, qui sont de véritables langues, parlées par des humains il y a quelques milliers d'années, le protolangage serait un système beaucoup plus rudimentaire, né il y a quelques centaines de milliers d'années. Selon Bickerton, il aurait possédé un lexique, mais pas de syntaxe, ou plus exactement une syntaxe très lâche, sans marques formelles : les phrases du protolangage auraient été composées de quelques mots lexicaux (verbes, noms, adjectifs) juxtaposés, sans ordre des mots bien défini, sans marques flexionnelles (pas de déclinaisons ni de conjugaisons) ni de mots grammaticaux (ni prépositions, ni conjonctions, ni déterminants). Un tel système de communication, consistant en une sorte de style télégraphique ou de langage « Tarzan », suffit, de fait, à échanger de l'information factuelle simple.

A l'appui de cette thèse du protolangage, Bickerton décrit un certain nombre de phénomènes, où l'on peut, selon lui, retrouver des « fossiles » de l'évolution du langage. Le premier phénomène concerne l'émergence des langues créoles, parlées notamment aux Antilles. Ces langues sont l'aboutissement d'une évolution en deux temps. D'abord les populations autochtones dominées, ou les esclaves dans le cas de la traite des Noirs, ont formé, à partir du lexique de la langue des maîtres (anglais, néerlandais, français, portugais ou espagnol, selon les cas) des pidgins, moyens de communication rudimentaires qui ne possédaient effectivement pas de syntaxe. Dans un deuxième temps, ces pidgins se sont transformés en créoles, véritables langues, devenues les langues maternelles des descendants des locuteurs de pidgins, et qui sont dotées de toutes les propriétés des langues humaines, y compris une syntaxe aussi complexe que celle des autres langues. Ce processus de création d'une langue en deux étapes serait analogue à celui du passage postulé du protolangage au langage. La deuxième série de faits qu'analyse Bickerton concerne les recherches sur des grands singes (notamment des chimpanzés et des bonobos) auxquels on a tenté d'enseigner une langue humaine, par diverses méthodes : apprentissage d'une langue des signes, communication à l'aide d'un clavier, etc. Ces tentatives ont abouti à des demi-succès : les

animaux arrivent à apprendre un vocabulaire non négligeable, et à utiliser des combinaisons de quelques signes de manière pertinente, mais ils semblent incapables de maîtriser la syntaxe de nos langues. Bickerton interprète ces résultats en avançant qu'ils ne pourraient donc pas dépasser le premier stade de l'évolution du système de communication humain, celui du protolangage. Il interprète de la même façon le cas d'enfants « sauvages » comme Genie, une enfant mutique de treize ans, maltraitée et cloîtrée depuis sa plus tendre enfance, qui une fois prise en charge et soignée, n'a effectivement jamais pu dépasser non plus le stade de production de phrases élémentaires sans structure syntaxique ni éléments grammaticaux. Bickerton tire aussi argument du déroulement normal de l'acquisition du langage par l'enfant. Il estime que l'on retrouve dans ce processus ontogénétique une « récapitulation » de l'histoire philogénétique : les enfants passeraient par une étape où leurs productions ressembleraient à du protolangage avant de maîtriser la syntaxe de leur langue.

Pour Bickerton, l'augmentation continue de la capacité crânienne des *Homo erectus* au cours de son évolution peut être mise en relation avec un accroissement de ses capacités cognitives, et notamment de ces capacités de catégorisation de son environnement, qui aurait produit un enrichissement progressif du lexique du protolangage. En revanche, le passage du protolangage au langage aurait été le fait de notre espèce, et se serait produit à capacité cervicale constante, si l'on peut dire, par une brusque réorganisation corticale qui aurait permis au système cognitif dédié à la communication d'utiliser un mécanisme déjà présent pour d'autres activités cognitives, lui conférant ainsi la compétence syntaxique nécessaire pour le traitement du langage humain.

Bien d'autres thèses ont été avancées, qui proposent elles aussi une évolution du langage en plusieurs étapes, sans adhérer pour autant à la description du protolangage proposée par Bickerton. Mentionnons simplement, à titre d'exemple, la théorie de Michael Corballis, pour qui le langage s'est d'abord développé chez *Homo erectus* sous forme gestuelle, se complexifiant graduellement jusqu'à devenir proche des langues des signes actuellement utilisées par les sourds, la dernière étape, franchie par *Homo sapiens*, étant le passage du geste à la parole. Signalons aussi la théorie de Merlin Donald, qui associe à *Homo erectus* un stade de l'évolution cognitive qu'il appelle la « culture mimétique », caractérisée par des capacités de représentation intentionnelle et de modélisation par le mime de l'expérience vécue, ce qui auraient permis la mise en place d'une communication fondée sur des jeux mimétiques réciproques et des actes mimétiques de groupe, bien adaptée à la vie sociale de plus en plus complexe de ces hominidés. Pour Donald, la dernière étape, franchie par *Homo sapiens*, aurait

conduit à ce qu'il appelle la « culture mythique », et elle aurait été rendue possible par l'acquisition d'une capacité nouvelle, celle d'inventer des symboles : cette théorie est donc diamétralement opposée à celle de Bickerton, puisqu'elle implique que le lexique aurait été mis en place lors de la dernière étape, alors qu'il aurait été présent dès le début du protolangage selon Bickerton.

Les causes de l'émergence du langage

Pourquoi le langage est-il émergé au cours de l'évolution des hominidés ? Cette question peut paraître saugrenue, tellement il est évident aujourd'hui que le langage constitue un avantage décisif pour notre espèce. Mais d'un point de vue darwinien, l'émergence d'un système de communication permettant d'échanger des informations factuelles sur le monde (sur les sources de nourriture, les prédateurs, etc.) pose au contraire un problème bien connu en théorie de l'évolution : celui de l'émergence de comportements altruistes. Un individu qui communique de l'information à ses congénères perd un atout susceptible de l'avantager dans la compétition pour la reproduction de son propre patrimoine génétique. Il n'y a donc aucune raison qu'un tel comportement se propage de génération en génération, puisque celui qui se conduirait ainsi serait systématiquement supplanté par ses congénères plus opportunistes, qui tireraient parti des informations fournies par les autres sans jamais en donner eux-mêmes, ou, pire, en divulguant de fausses nouvelles. Autrement dit, la sélection naturelle devrait privilégier des individus mutiques ou menteurs, ce qui fait qu'en retour plus personne n'aurait intérêt à écouter. On conviendra qu'un système de communication aurait du mal à se maintenir dans ces conditions !

Plusieurs hypothèses ont été formulées pour résoudre ce paradoxe, qui, il faut le souligner, se pose dès l'apparition du protolangage, du moins si l'on accepte avec Bickerton que le protolangage servait à communiquer de l'information factuelle. Parmi les solutions proposées, on peut citer celle de Jean-Louis Dessalles, qui invoque l'aspiration à s'élever dans la hiérarchie sociale qui devait caractériser l'organisation des sociétés d'hominidés. En donnant de l'information à ses congénères, un individu obtiendrait en échange un meilleur statut social. La qualité de l'information transmise aurait ainsi remplacé en grande partie, chez nos lointains ancêtres, la force physique et autres qualités régissant, chez nos cousins chimpanzés, la position sociale de chacun, la constitution de coalitions, etc. On peut aussi mentionner la thèse de Robin Dunbar, intitulée « *grooming and gossip* », qui défend l'idée que la communication orale a remplacé avantageusement les séances de toilettage (*grooming*) qui

permettent de nouer et de maintenir les alliances entre congénères dans les sociétés de primates. L'augmentation de la taille des groupes sociaux aurait rendu le toilettage trop coûteux en temps, et, de plus l'échange d'information sur le comportement des congénères (*gossip*) présente l'avantage de mieux détecter les tricheurs, qui chercheraient à tromper leurs partenaires sur leurs intentions réelles. Dans ces deux thèses, on le voit, l'apparition d'un système d'échange d'information serait donc à mettre en relation avec d'autres spécificités des hominidés, comme le développement de l'intelligence et la complexification des rapports sociaux.

Un deuxième problème concerne plus précisément la dernière étape de l'évolution du langage : pourquoi nos langues sont-elles si complexes, autrement dit pourquoi sont apparues des innovations comme les propriétés syntaxiques qui confère au langage humain une puissance expressive sans aucune commune mesure avec le protolangage qui l'aurait précédé ? Ce qu'il faut comprendre, c'est quel type d'information réclame cette complexité, et surtout quel type d'information a pu nécessiter un tel degré de complexité dans le passé. A-t-elle été rendue indispensable par la mise en place d'activités humaines complexes comme la chasse ou la fabrication d'outils ? Le fait que les chimpanzés soient tout à fait capables de chasser en groupe et de fabriquer des outils sans avoir eu besoin pour autant de développer un langage similaire au nôtre ne milite pas en faveur de cette hypothèse. De plus, un protolangage semble suffisant pour conduire une chasse ou pour enseigner une activité manuelle. Il faut donc plutôt se pencher sur des fonctions qui ne peuvent pas être accomplies en l'absence d'une syntaxe élaborée. Jean-Louis Dessalles, par exemple, met de l'avant la fonction argumentative du langage : elle repose en effet sur des raisonnements, dont l'expression réclame de manière cruciale des outils grammaticaux spécifiques (comme les propositions conditionnelles), et elle permet de mieux contrôler la fiabilité de l'information donnée par un locuteur en s'assurant de son bien-fondé. D'autres, comme Mark Turner, penchent plutôt pour la fonction narrative : toute la complexité grammaticale de la langue est en effet à l'œuvre dans la narration, et c'est une activité universellement répandue dans les sociétés humaines, dont le rôle dans l'apprentissage et la régulation des comportements sociaux est essentiel. Mais il faut reconnaître que ces différentes thèses restent hautement spéculatives.

Ainsi, l'origine du langage et des langues reste encore énigmatique par bien des aspects, bien que nos connaissances sur l'évolution des hominidés et sur l'histoire de notre espèce aient progressé de manière très significative au cours des dernières décennies. Mais si cette

question fondamentale, qui a toujours fasciné les hommes, n'est pas prête d'être définitivement résolue, il faut souligner qu'elle devient scientifiquement productive, au sens où elle se pose aujourd'hui en des termes dont l'étude fait progresser notre compréhension de l'homme et de sa place dans le monde vivant.

Bibliographie

- Bickerton D., *Language and Species*, University of Chicago Press, 1990.
- Cavalli-Sforza L., *Gènes, peuples et langues*, Odile Jacob, 1996.
- Corballis M.C., *From Hand to Mouth*, Princeton University Press, 2002.
- Dessalles J.-L., *Aux origines du langage – une histoire naturelle de la parole*, Hermès, 2000.
- Donald M., *Origins of Modern Mind: Three Stages in the Evolution of culture and Cognition*, Harvard University Press, 1991. (*Les origines de l'esprit moderne*, DeBoeck Université, 1999).
- Dunbar R., *Grooming, Gossip and the Evolution of Language*. Faber Faber and Harvard University Press, 1996.
- Hombert J.M. (éd.), *Aux origines des langues et du langage*, Fayard, 2005.
- Johansson S., *Origins of language, constraints on hypotheses*, John Benjamins, 2005.
- Laks B., Victorri B., édés., *L'origine du langage*, n° 146 de la revue *Langages*, Larousse, 2002.
- Lieberman P., *Uniquely Human: The evolution of speech, thought and selfless behavior*. Harvard University Press, 1990.
- Pinker S., *The Language Instinct*, William Morrow and Co, 1994 (*L'instinct du langage*, Odile Jacob, 1999).
- Ruhlen M., *The Origin of Language : Tracing the Evolution of the Mother Tongue*, John Wiley & Sons, 1994 (*L'Origine des langues. Sur les traces de la langue mère*. Belin, 1997).
- Tattersall I., *Becoming Human, Evolution and Human Uniqueness*, Harcourt Brace and Co., 1998.
- Turner M., *The Literary Mind*, Oxford University Press, 1996.