

***Journées du Patrimoine 2003
vendredi 19 septembre, 20h30***

***Bayonne
Auditorium C.C.I.***

***Conférence :
« Langue et peuplement : le cas de la langue basque »***

Intervention de Jon LANDABURU, directeur du Centre d'études des langues indigènes d'Amérique du CNRS

*Gau on, jaun-andereak, euskaldunak, euskaltzaleak, guzieri agur bero bat...
Ez nuke gaur mahai honetan egon behar, nere jakintza -eskasa-
eta nere ardurak -handiak- hogeita hamar urte hauetan Ameriketako hizkuntz gaietan izan
baitira. Baina hemen dituztan lagun zaharrak, Beñat Oyharçabal eta Jean-Marie Hombert,
tematu dira ni beren artean izaiteko gaur, eztabaida honetan nere esperentzia sar nezan.
Azkenean, nola ez onartu holako gonbidapena ?*

*Hasi baino lehenago, lagun haueri eta Euskal Kultur Erakundeari, eskerrik asko.
Horra hemen, gogoeta batzuk, frantsezez, hizkuntzen jatorriari buruz... Hea argi apur bat edo
nahaste-borraste gehiago ekarriko dizuedan...*

Je disais en euskara que je ne sais pas si je suis vraiment au bon endroit puisque je ne suis pas un bascologue professionnel mais seulement un linguiste basque (et non du basque). Je suis en fait spécialiste des langues amérindiennes de l'Amérique du Sud et plus particulièrement de la Colombie que je pratique depuis maintenant plus de trente années. C'est peut-être à ce dernier titre que l'on m'a fait l'honneur d'être invité à cette table-ronde, puisque les problèmes de classification des langues et des peuples se posent également en Amérique indienne et l'évaluation des nouvelles méthodes et des nouvelles hypothèses en la matière y a aussi produit d'intenses discussions. Mon propos sera d'illustrer quelques unes des avancées et des difficultés apportées par ces nouveaux apports, tant à un niveau général qu'au niveau plus restreint de la problématique amérindienne, en laissant à chacun la charge d'en tirer les leçons qu'il souhaitera pour le cas particulier de la langue basque.

Classer une langue c'est la rapprocher d'autres langues en essayant d'établir un réseau de relations de proximité entre toutes ces langues en fonction des signes qu'elles partagent. On peut dire qu'un signe est partagé entre deux (ou plusieurs) langues, si un mot, ou une expression verbale, est aussi utilisé par quelqu'un qui parle une autre langue, dans cette autre langue. Le signe ne sera jamais exactement le même d'une langue à l'autre mais on gardera l'idée de partage si d'un signe à l'autre se maintient un rapport de proximité phonétique lié à un rapport de proximité sémantique. Par exemple le français et le castillan partagent un signe sous la forme de *plaga* en castillan et de *plaie* en français. Même si la forme phonétique et le sens ne sont pas exactement les mêmes, on peut percevoir une proximité entre les deux mots

du côté phonétique comme du côté sémantique. On voit par cet exemple que ce n'est pas exactement un signe qui est partagé mais plutôt un « instrument à signifier », une espèce d'entité relationnelle bizarre cachée qui évolue en prenant des formes et des sens différents dans des parlers ou des langues différentes. Si un nombre important de ces « instruments de signification » sont partagés entre deux langues, ce qui est un fait d'observation courante, par exemple entre le français et le castillan, on se pose la question de la cause de ce partage massif. Une des explications habituelles est que les deux langues procèdent de l'évolution d'une seule langue selon deux traditions qui se sont séparées.

Dans cette optique, les linguistes ont pris l'habitude de parler de familles de langues, de langues mères ou matrices, de langues filles, etc., et ont introduit, ou plutôt repris il y a déjà longtemps, un vocabulaire de parenté dans un domaine où il n'est peut-être pas approprié. Classifier des langues est ainsi normalement vu comme un essai pour tracer un réseau de langues « parentes » plus ou moins lointaines¹. On ne s'étonnera pas qu'une telle tentative généalogique ou génétique ait historiquement soulevé des émotions et des intérêts idéologiques qui interfèrent avec une approche sereine d'un problème scientifique.

Du côté des locuteurs d'une langue, la recherche d'une parenté linguistique soulève évidemment des questions d'identité. Le parent c'est d'abord celui qui m'engendre (du latin *parere* « engendrer, produire », d'où le castillan *parir* « accoucher ») et par extension celui qui est lié à moi à travers des rapports d'engendrement, des mariages ou des filiations. Mon identité est comme ma chair, elle est un corps symbolique qui me définit et trace un espace où s'opposent le soi et l'autre : les gens qui sont comme moi, face à ceux qui ne sont pas comme moi. Comme je ne me pose pas moi-même dans l'existence mais que je suis engendré, mon identité dépend pour une bonne part de mes parents. Parents d'abord au sens restreint : père et mère ; mais aussi parents au sens large, à différents degrés de proximité. Normalement, on ne choisit pas ses parents, mais quand les choses ne sont pas très sûres, on peut jouer à s'en donner, selon le goût du jour. On peut tout aussi bien se les refuser pour profiter du prestige inverse de l'originalité, de la singularité. On bénéficie alors du mystère qu'a celui dont on ne sait pas d'où il vient. Le même jeu se produit avec la langue, et l'établissement du soi et de l'autre à divers degrés donne lieu à beaucoup d'intérêts idéologiques, car on passe facilement d'une classification des langues à une classification des communautés linguistiques. On passe des signes qui sont partagés aux gens qui utilisent ces signes. On passe rapidement, trop rapidement, de la parenté des langues à la parenté des peuples.

Les linguistes spécialisés ont souvent du mal devant certaines inévitables demandes et n'arrivent la plupart du temps pas à s'en dépêtrer. Conscients que les résultats de la recherche scientifique sont fragiles et relatifs, et soucieux de ne pas être utilisés dans des enjeux qui ne les concernent pas en tant que linguistes, ils préfèrent souvent s'abstenir et en rester à une position de non intervention voire, ce qui est moins honorable, à une position de non recherche. Etant donné l'état de la recherche on décide que tout cela est beaucoup trop obscur, et on ne se lance pas dans les questions de parenté ou d'origine.

Le succès des récentes publications sur l'origine des langues, et sur l'origine du basque en particulier, s'appuie directement sur cette situation et les affirmations démagogiques d'un Merritt Ruhlen sur l'obscurantisme des linguistes qui s'opposent à ces

¹ On parle alors précisément de classification « génétique » et on l'oppose aux classifications « typologiques » qui concernent des langues qui partagent non pas des signes mais des traits. Les langues peuvent être en effet aussi rapprochées parce qu'elles partagent des traits phonétiques (par exemple celles qui ont des tons) ou des traits sémantiques (par exemple celles qui associent le passé au présent devant le futur face à celles qui associent le passé au futur devant le présent), ou des traits syntaxiques qui sont les modes d'organisation des signes (par exemple celles qui tendent à mettre l'objet devant le verbe face à celles qui tendent à le mettre après, etc.). Le rapprochement entre traits peut être interprété comme provenant d'un contact prolongé et d'un emprunt, ou comme le fruit d'une coïncidence, les langues ayant alors chacune inventé de son côté le trait partagé ; il n'implique pas nécessairement une commune origine.

théories y trouvent un terrain favorable. Il faut reconnaître à Joseph Greenberg, Cavalli-Sforza et aux partisans de ce qui est assez pompeusement appelé *emerging synthesis* le mérite d'avoir relancé le débat et secoué la profession. Ils n'ont pas été les seuls, ni mêmes les premiers à être récemment revenus sur ces problèmes d'origine et de superfamilles de langues, mais la passion mise dans les affirmations, les argumentations, les attaques, la reprise des discussions dans les grands moyens de communication, ont réveillé beaucoup de linguistes qui ne s'intéressaient pas ou plus à ces problèmes, problèmes qui sont pourtant réels et importants même si on avait désespéré de pouvoir un jour les aborder.

Qu'en est-il de ces avancées supposées et sur quoi s'appuient-elles ? Face à la tradition de la linguistique comparative historique que Greenberg considère à juste titre comme incapable de remonter profondément dans le temps et qui n'avait effectivement pas pu résoudre le problème des langues dites isolées comme le basque ou comme de nombreuses langues indiennes d'Amérique, ce linguiste nord-américain propose de nouvelles hypothèses de regroupements à partir de deux méthodes nouvelles :

- 1) une méthode dite de comparaison multilatérale qui interprète la répartition dans les langues du monde de signes ressemblants en classes de proximité entre langues, et ceci, qu'on ait trouvé ou non des lois de correspondances systématiques entre les sons de ces signes (seules ces lois de correspondance entre sons autorisaient traditionnellement les linguistes à poser des parentés),
- 2) une méthode fondée sur les correspondances qui s'établissent entre les classes de proximité linguistique issues de la première méthode, et les classes de proximité qu'établissent maintenant les généticiens entre structures génétiques des populations. Ces correspondances permettraient aux linguistes de poser des apparentements entre langues sur la base de données génétiques, même si les évidences linguistiques sont insuffisantes.

Les résultats issus de ces deux méthodes sont considérables puisqu'ils auraient permis à Greenberg et surtout à ses disciples (Ruhlen, Bengston, etc.) de poser, à partir d'une souche africaine, l'unicité originelle de toutes les langues humaines actuellement existantes, d'affirmer l'unicité de toutes les langues amérindiennes (à l'exception des familles Eskimo-aléoute et Na-dene), de proposer un schéma de migration des peuples depuis l'Afrique d'il y a 100.000 ans, de poser des liens entre l'euskara et une famille Caucase-Na dene, etc. Le débat de ce soir n'est évidemment pas le lieu pour entrer dans les détails et les difficultés techniques présentes dans chacune de ces deux démarches, ou dans l'analyse de la justesse des conclusions qui en sont tirées. Nous nous bornerons à signaler quelques problèmes, suffisamment graves, pensons-nous, pour nous obliger à demeurer circonspects et attendre de nouvelles avancées.

Le premier problème que nous évoquerons provient de la difficulté de tirer des conclusions sur l'apparentement des langues à partir des faits de distribution de traits génétiques. Un même « scénario » génétique peut en effet être compatible avec des scénarios linguistiques complètement différents. Prenons précisément un de ces récents articles qui ont suscité un regain d'intérêt pour le problème de l'origine des Basques. Dans « L'épopée du génome basque » paru dans *Pour la Science* il y a un an (N°299 : 33), Elizabeth Hamel et Peter Forster nous présentent la situation des populations de l'Asie de l'est. Ils disent :

« Nous y avons observé un échelonnement génétique entre le nord et le sud, avec une frontière située à la latitude de la Corée. Au sud de cette limite, nous avons recensé six types d'ADN mitochondrial âgés de plus de 30.000 ans ; au nord, les lignées sont toutes des descendantes de celles du Sud et ont toutes moins de 16.000 ans. Ainsi les hommes se sont réfugiés dans les régions chaudes du sud pendant la période glaciaire, puis sont retournés peupler le Nord, lorsque le climat est redevenu clément. Ils s'y seraient installés,

effaçant les traces qui subsistaient encore des ancêtres des Indiens (américains) en Sibérie orientale. »

Quelles conclusions tirer de cette situation pour les langues ? L'auteur nous dit lui-même avec sagesse: « Nous ignorons l'équivalent linguistique de ce scénario génétique », mais cela ne l'a pas empêché d'en proposer un pour l'Amérique quelques lignes avant, en fait uniquement parce que le linguiste Joseph Greenberg en avait proposé un. Les généticiens s'appuient sur les « faits » du linguiste ; les linguistes, au moins les disciples de Greenberg, s'appuient sur les « faits » du généticien. Qu'en est-il ? Nous ne saurions prétendre mesurer le degré de validité des faits du généticien et il faut en laisser la discussion à des spécialistes. Les résultats ne sont peut-être pas en tout cas aussi assurés et les problèmes de taille d'échantillonnage ou ceux du suivi du chromosome Y face au suivi de l'ADN mitochondrial, semblent redoutables. Pour le cas qui nous intéresse, dans notre perspective de linguiste, pour un même scénario génétique, plusieurs scénarios linguistiques semblent possibles à partir de la donnée d'un repeuplement, il y a ± 16.000 ans, de l'Asie du Nord par l'Asie du Sud :

- 1) Les nouvelles populations venant du sud gardent leurs langues et aucun contact n'a lieu avec les langues du fonds archaïque sibérien. Ces dernières disparaissent ou ont déjà disparu. Il y a cohérence entre l'hypothèse de peuplement issue des gènes et le regroupement des langues.
- 2) Il y a encore des populations autochtones et les nouvelles populations, sans se mélanger aux autochtones de manière massive, sont en contact et acceptent des emprunts lexicaux voire structurels. Il y a alors présence, dans les langues, d'éléments qui renvoient à d'autres populations que celles attestées par les gènes. Peut-être est-ce l'origine de ressemblances structurelles entre les langues sibériennes actuelles et certaines langues amérindiennes. Il n'y a pas de correspondance globale claire entre traits génétiques et signes linguistiques.
- 3) Les nouvelles populations, selon différents scénarios possibles (degrés très divers de métissage), prennent la langue des autochtones. Il existe des exemples de cette situation comme le cas des Garifuna de la côte atlantique d'Amérique centrale (Honduras, Belize, Guatemala) provenant d'esclaves africains ayant complètement adopté une langue amérindienne parlée par leurs hôtes arawako-caribes ou les Puinave du Nord-Ouest amazonien qui sont des populations horticoles parlant une langue proche de celle des Maku, chasseurs-cueilleurs de la région. On a aussi de bons documents (travaux de Robert Nicolai) sur la langue songhay septentrionale du Niger parlée par des populations nomades de type touareg et dont le stock lexical est à moitié d'origine touareg, à moitié d'origine négro-africain. Les hypothèses de peuplement qu'autoriseraient dans ces cas les données génétiques sont alors plutôt en contradiction avec celles qu'induiraient les données linguistiques.

Les scénarios linguistiques possibles pour des situations où il y a eu brassage important de populations provenant de souches distinctes sont encore plus variés et complexes. Et on ne voit pas comment la distribution des traits génétiques pourrait nous aider à reconstituer l'histoire des langues en contact, l'histoire des langues mixtes, l'histoire des pidgins et des créoles, etc. Comment peut-on exclure de telles situations pour le passé, même pour un lointain passé ?

Notre réticence à accepter des inférences directes du génétique au linguistique provient *in fine* de la différence fondamentale entre la transmission héréditaire de gènes par laquelle un organisme se construit à partir d'organismes antérieurs –ordre de la génétique- et la transmission d'habitudes ou de normes, que sont les langues –ordre de la linguistique. Les langues ne sont pas des individus de type biologique ; ce sont des systèmes de signes, des institutions sociales au sens où elles sont un corps organisé de pratiques et de valeurs. Elles ne sont pas des êtres vivants qui se prolongent en d'autres êtres. Elles supposent des êtres vivants

qui les parlent et qui les transmettent. Elles ne naissent ni ne meurent qu'à travers des individus, elles ne se reproduisent pas dans d'autres langues.

Leur mode d'être, car elles correspondent quand même à un type d'objet, peut être en partie décrit par le terme de « tradition » et nous entendons par là un système d'habitudes qui se transmet, se modifie et s'adapte. Car les langues changent, et en ce sens on peut bien parler d' « histoire des langues » mais, contrairement aux génomes qui sont soumis à des mutations apparemment erratiques, les langues, par delà leur inertie propre qui induit des changements mécaniques, subissent aussi des changements fonctionnels en attrapant des éléments d'autres langues, des mots mais aussi des procédés phonétiques, des mécanismes grammaticaux, des trouvailles sémantiques. Nous avons tort en fait de dire que les langues attrapent des éléments d'autres langues. Ce sont toujours des individus humains réels qui le font et ce point, trop longtemps occulté par la linguistique des XIX^{ème} et XX^{ème} siècles est en train d'apparaître davantage dans les recherches actuelles sur le changement linguistique. Pour une bonne part les langues sont des constructions sociales soumises à l'innovation des individus et non des destins. Les populations ne choisissent pas leur langue de départ mais elles peuvent en changer et la changer. Ce qui est retenu ne dépend pas de l'individu mais ce qui change a été inventé par des individus dans un environnement social, culturel et technologique particulier.

Plus qu'une correspondance factuelle entre langues et gènes opérée par des généticiens parfois trop peu sensibles à la complexité et à la multidimensionnalité des langues, les linguistes devraient par contre retenir de la génétique son concept de population et son traitement statistique de l'innovation génétique. De même que les généticiens nous ont obligé à revoir l'idée d'espèce, c'est le concept de langue, monolithique et essentialiste qui doit être revu par les linguistes. Il l'est en fait déjà mais sa version ancienne réapparaît souvent lorsqu'on parle de familles linguistiques et de classification. Réintroduire les acquis de la discipline dans ces discussions est une tâche difficile qui permettrait cependant de sortir des débats trop influencés par les questions identitaires.

La seconde série de problèmes provient de la méthode de comparaison multilatérale, qu'on l'examine en elle-même ou dans la manière dont elle a été appliquée. Rappelons que par cette méthode (qui est en fait plus une technique qu'une méthode) les familles de langues sont constituées en fonction du nombre de ressemblances trouvées entre langues. On dresse la liste d'un petit nombre de notions telles qu'elles apparaissent dans un grand nombre de langues et on repère directement les formes ressemblantes d'une langue à l'autre. On établit des classes de langues plus ou moins proches selon le nombre de ressemblances trouvées (c'est là qu'on se sépare de la méthode traditionnelle qui va travailler sur les premières ressemblances trouvées, en induire des hypothèses de relations systématiques entre formes phonétiques, associer à partir de là de nouvelles paires non visibles à l'observation naïve, en rejeter d'autres, etc.) et on suppose que l'appartenance à une classe implique une commune origine linguistique et, de là, une commune origine génétique. On organise ces classes les unes par rapport aux autres dans un schéma arborescent qui donne la généalogie des langues. Cette méthode n'est pas nouvelle et des chercheurs plus anciens comme Kroeber ou Paul Rivet l'avaient utilisée. Le problème principal réside dans la nature de la ressemblance trouvée. La ressemblance peut en effet renvoyer à une origine commune ou à bien d'autres raisons. Greenberg et ses disciples ne s'en soucient pas et considèrent que statistiquement les erreurs d'attribution possibles jouent dans tous les sens et finissent par se neutraliser.

Cela n'est pas satisfaisant pour la plupart des linguistes qui pensent que, même si la rigidité de la méthode comparative traditionnelle (qui ne garde les formes ressemblantes que si on a établi des règles de correspondance systématique entre sons) est probablement excessive, on doit rejeter - car ne relevant pas d'une commune origine- les formes dont la ressemblance provient de l'emprunt, du hasard, de l'onomatopée ou du symbolisme sonore. Ainsi beaucoup de langues utilisent une forme comme « mama, ama, nana, ana, etc.. » pour

désigner la mère et il semble bien que l'utilisation par l'enfant d'un patron bisyllabique autour d'une nasale labiale ou dentale relève davantage de mécanismes psycho-phonétiques fondamentaux et toujours accessibles que de la transmission d'une forme apprise. Même si chaque langue à ses onomatopées (français *cocorico*, basque *kukurruka*), on voit bien que l'harmonie imitative impose des solutions qui peuvent être proches d'une langue à l'autre sans que les deux formes soient issues nécessairement l'une de l'autre. Ainsi plusieurs langues amazoniennes désignent la flûte par un mot de type *tchiru*, *fidu*, etc. qui rappelle le basque *txirula*. Enfin l'emprunt est la principale difficulté que rencontre cette méthode d'observation directe car il est souvent difficile de décider, lorsqu'il n'y a pas de témoignages écrits, que la cause de la ressemblance entre deux mots de deux langues proches géographiquement est une forme « ancestrale », ou l'emprunt de l'une à l'autre, ou l'emprunt des deux à une troisième apparentée ou non. Comme le dit un important linguiste amérindianiste (Marianne Mithun 1999, *The Languages of Native North America* : 311) :

« Les langues en viennent souvent à partager leurs caractéristiques entre leurs locuteurs. Aux premières étapes du contact, des noms sont partagés pour désigner des choses nouvelles comme par exemple des marchandises ou des noms de lieux. Avec un contact plus intime et spécialement avec le bilinguisme, l'emprunt peut se généraliser et inclure non seulement une vaste part du vocabulaire mais aussi des traits phonologiques, grammaticaux et sémantiques... Avec le temps, un groupe de langues peuvent partager un nombre important d'éléments. Un tel groupe est appelé « Sprachbund » et la région géographique dans laquelle ces emprunts importants ont eu lieu s'appelle « aire linguistique » ou « aire de convergence » ou « aire de diffusion ».

De telles aires de convergence ont dû se produire à plusieurs reprises lors de l'exploitation par plusieurs groupes d'origine différente d'un même écosystème, sans qu'on puisse aisément déterminer pour chacun ce qui relève dans la langue d'une origine commune et ce qui relève de l'emprunt.

Outre ces difficultés de fond, et pour nous cantonner au champ des études sur les langues indiennes d'Amérique, beaucoup de linguistes amérindianistes reprochent aussi à Greenberg et à ses disciples de ne pas être assez précautionneux dans l'application même de la méthode. Les spécialistes de langues ou de familles américaines particulières ont très souvent constaté une utilisation de données de seconde main peu fiables (cela a été aussi mon cas pour les langues que je connais et pour lesquelles j'ai des données de première main), une reconnaissance de classifications génétiques antérieures faites par des auteurs plus anciens acceptées sans bénéfice d'inventaire alors qu'elles sont contestées par les spécialistes, la comptabilisation pour la comparaison de formes monosyllabiques trop courtes pouvant de ce fait renvoyer à des coïncidences fortuites, l'utilisation de ressemblances accidentelles dont on avait pourtant montré le caractère fallacieux (cas de la similitude des différentes formes de première personne en quechua et en quiché de la famille Maya), une analyse erronée des mots permettant des comparaisons fallacieuses entre morphèmes inexistantes, etc. Toutes ces difficultés ont largement contribué à frapper de soupçon les hypothèses et les méthodes nouvelles auprès des spécialistes des langues américaines. Ce que nous avons su des critiques de Trask sur les affirmations et la faiblesse des données de Bengston concernant l'inclusion de l'euskara dans une famille Dene-caucasienne avec le buruchaski, le ienissei et la famille sino-tibétaine vont dans le même sens.

L'utilisation par Merritt Ruhlen du fait que la classification des langues africaines de Greenberg ait été généralement acceptée, comme un argument pour renforcer l'autorité de sa classification des langues amérindiennes est de plus un argument spécieux. En effet alors que la plupart des spécialistes des langues africaines acceptent la classification de Greenberg, la plupart des spécialistes de la comparaison des langues amérindiennes ne l'acceptent précisément pas. Le succès de sa classification des langues d'Afrique est certes un préjugé

favorable permettant de regarder avec attention et considération sa classification des langues d'Amérique. Mais si on retient dans un cas l'avis des spécialistes pour prendre au sérieux une classification, on ne saurait dans un autre cas disqualifier cet avis. Comme le dit Lyle Campbell, un des plus durs critiques de Greenberg : « Le succès ailleurs ne garantit pas le succès ici ».

En conclusion, nous pourrions dire que les spectaculaires avancées qu'on nous promettait en ce qui concerne la classification des langues sont encore à venir, qu'indiscutablement les progrès de la génétique dans l'établissement de lignées humaines semblent pouvoir nous préciser un cadre dans lequel il faudra essayer de tester plusieurs scénarios linguistiques, que la secousse apportée par ces débats à la linguistique historique a contribué à reposer des problèmes anciens qu'on ne sait toujours pas résoudre, mais que cette même conscience des difficultés peut contribuer à inventer de nouvelles manières de penser l'évolution des langues.