

**La classification des langues en familles et macro-familles,
avec une attention particulière portée aux
langues d'Asie Orientale et d'Asie du Sud-Est**

**Alain Peyraube
CRLAO (CNRS et EHESS)**

Familles et macro-familles

Les linguistes regroupent traditionnellement les 5000 à 6000 langues du monde – dont près de la moitié disparaîtra au cours du siècle – dans 400 à 500 familles, de taille très inégale¹. Certaines d'entre elles, comme la famille austronésienne, comptent plus de 1200 langues, d'autres n'en comptent qu'une seule : le basque, par exemple, exemple bien connu d'isolat linguistique. L'Europe, quant à elle, est plutôt pauvre. L'ALE (*Atlas linguarum Europae*) reconnaît officiellement 88 langues, regroupées en 17 branches constituant six familles (*phyla*), dont l'indo-européen (IE), famille de loin la plus importante (12 branches selon Contini 2000).

Les linguistes ont plutôt eu pour habitude, jusqu'à présent, d'essayer de reconstruire des proto-langues (dont les plus anciennes ne remonteraient pas au-delà de 6000 ans avant J.-C.) pour les familles de langues dont ils sont spécialistes. Rares sont ceux qui ont tenté, dans le passé, de comparer entre elles des familles différentes. Ce parti pris méthodologique est en train d'évoluer.

Certes, des hypothèses de grands regroupements ont été régulièrement avancées tout au long du 20^{ème} siècle, mais elles n'ont jamais été véritablement prises au sérieux, comme celle de Trombetti (1905) qui pensait que toutes les langues devaient appartenir à une seule et même famille.

Au début du 20^{ème} siècle également, le linguiste danois Pedersen (1903) suggérait déjà de rattacher l'indo-européen (IE) à d'autres familles telles que l'ouralien (langues finno-ougriennes et samoyèdes), l'altaïque, l'esquimo-aléoute et le sémitique, au sein d'une macro-famille qu'il appela le nostratique. Il fallut cependant attendre le début des années 1960 pour que cette proposition soit sérieusement reprise et développée par l'école linguistique russe (Dolgopolski et Illitch-Svitytch), qui fit aussi entrer le kartvélien, le dravidien et le tchouktchikamtchatkien dans le nostratique. Cette thèse du nostratique ne s'est quand même jamais imposée dans la communauté linguistique internationale.

Il en a été de même pour l'austro-tai, qui regroupait à l'origine (au début du 20^{ème} siècle) l'austronésien et les langues môn-khmer ou austro-asiatiques. L'existence d'une famille austro-tai est toujours controversée, de même que la proposition de Benedict (1942) de rassembler le tai-kadai et l'austronésien dans une famille austro-tai.

La situation s'est quelque peu modifiée au début des années 1980. La proposition de Greenberg (1963a) de réduire la diversité des langues africaines à quatre macro-familles a été acceptée par une grande majorité de spécialistes, après avoir été l'objet de critiques virulentes de la part des africanistes. C'est aussi l'époque où les linguistes russes Starostin et Nikolaïev (Starostin 1989) présentent des faits qui militent en faveur d'apparentements entre le proto-caucasien, le proto-sino-

tibétain et le proto-iénisséen, dans un premier temps (hypothèse du sino-caucasien), avant que soit reliée, dans un deuxième temps, la famille caucasienne et la famille na-déné, qui avait été découverte par Sapir, mais à laquelle personne n'avait auparavant prêté une grande attention (Ruhlen 1994 : macro-famille déné-caucasienne). A la fin des années 1980, le linguiste américain Bengston (1991) est aussi convaincu que le basque et le bouroushaski, ainsi que le sumérien, pourraient être affiliés au déné-caucasien.

En 1987 enfin, Greenberg réduit les quelques deux cents familles indépendantes des Amériques (dont une soixantaine pour les langues d'Amérique du Nord, groupant 213 langues, selon Chafe 1962) à seulement trois familles, l'amérinde, l'althabasque et l'eskimo-aléoute, les deux dernières appartenant, qui plus est, à de plus grands ensembles. Plus tard, Greenberg pense aussi, comme les linguistes russes partisans du nostratique, que l'IE est clairement apparenté à d'autres familles d'Eurasie, comme l'ouralien, l'altaïque et l'eskimo-aléoute, et il parle d'une macro-famille eurasiatique (Greenberg, à paraître).

Ces hypothèses sont aujourd'hui abondamment discutées, notamment par les typologues dont le souci essentiel n'est pas de reconstruire des proto-langues, mais de proposer des classifications de langues. Ils pensent en effet qu'il n'est pas nécessaire que les familles soient d'abord reconstruites sous la forme de proto-langues pour qu'on puisse ensuite les comparer.

En retenant ces différentes propositions, Ruhlen dégage dès 1992 une douzaine de macro-familles qui engloberaient les 5000 à 6000 langues du monde. Cette réduction drastique du nombre de *phyla* remet assurément à l'ordre du jour le problème de la monogénèse des langues. Les douze macro-familles de Ruhlen (1992, 1994, 1997) sont les suivantes : Khoisan, Nilo-saharien, Nigéro-Kordofanien, Afro-asiatique, Kartvélien, Dravidien, Eurasiatique, Déné-caucasien, Austrique, Indo-Pacifique, Australien, Amérinde. Elles sont ainsi réparties par grandes zones géographiques :

Afrique

- Khoisan (Afrique du Sud, Tanzanie)
- Nigéro-kordofanien, qui occupe une vaste zone de l'Afrique subsaharienne et qui est composé de deux branches majeures : le kordofanien (sud du Soudan) et le nigéro-congolais (avec les centaines de langues bantoues, dont plus de quatre cents au seul Nigéria : zoulou, swahili, mbundu, etc.)
- Nilo-saharien, un ensemble de langues parlées dans le nord de l'Afrique centrale et en Afrique de l'Est, avec quelques dizaines de sous-familles, dont le nilotique.
- Afro-asiatique, qui comprend le sémitique (arabe, hébreu), le tchadique (haoussa, bole, ngamo, etc.), le berbère, l'ancien égyptien, l'omotique (kafa, mocha, dime), le couchitique (afar, somali).

Asie du Sud-Est et Océanie

L'Austrique est la seule macro-famille présente en Asie du Sud-Est. Elle regroupe l'austro-asiatique (composé des langues munda du nord de l'Inde et des langues môn-khmer du Vietnam et du Cambodge), le miao-yao (dans le sud de la Chine et au Vietnam), le tai-kadai (en Thaïlande et au Laos), l'austronésien (Taiwan, Malaisie, Indonésie où on recense 670 langues, Philippines, Madagascar, Nouvelle-Zélande, Tahiti, etc.).

Trois macro-familles sont réparties dans le continent océanien : l'austronésien (sous-famille appartenant à l'Austrique, voir ci-dessus), l'Indo-Pacifique (en

Papouasie-Nouvelle-Guinée, qui compte près de 800 langues) et l'Australien (plus de 200 langues).

Amériques

Trois familles seulement regroupent toutes les langues des Amériques :

- eskimo-aléoute (en Alaska ; il s'agit d'une sous-famille de l'Eurasiatique)
- na-déné (une sous-famille du Déné-caucasien qui rassemble les langues de la famille althabasque et d'autres langues de la côte méridionale d'Alaska)
- Amérinde, une macro-famille divisée en onze sous-familles et qui comprend, entre autres : en Amérique du Nord et centrale, l'almosan, le kere'sion, le penutia, le hokan, l'algonquin, l'uto-aztèque, l'amérinde central, etc. ; en Amérique du Sud, l'andin, l'arawak, le macro-tucano, le macro-caribe, le macro-pano, le macro-ge, etc. Plus de deux cents langues indiennes sont parlées au seul Brésil.

Eurasie

- Dravidien (Inde du sud) : tamil, brahoui
- Kartvélien (Géorgie)
- Eurasiatique, qui réunit l'indo-européen (anatolien, langues romanes [français, espagnol, portugais, italien, roumain, français, auxquelles on peut rajouter l'occitan, le catalan, le gallicien, le rhéto-romanche, le corse et le sarde], langues germaniques, tokarien, etc.), l'ouralien (quelques vingt-cinq langues finno-ougriennes [finnois, hongrois, lappon, estonien], ainsi que les langues samoyèdes), l'altaïque (divisé en trois branches turque, mongole et tongouso-mandchoue), un groupe coréen-japonais-aïnou, le tchouktchi-kamtchatkien (en Sibérie du Nord et Orientale), l'eskimo-aléoute (groënlandais)².
- Déné-caucasien, qui comprend : le basque (Pyrénées), le caucasien (dont le tchéchène), le bouroushaski (montagnes du Nord-Pakistan), le iénisséen (le ket parlé en Sibérie centrale), le sino-tibétain, le na-déné.

Cette classification est bien loin d'être l'objet d'un consensus, même vague, de la part des linguistes. Si les quatre macro-familles des langues africaines de Greenberg sont aujourd'hui acceptées par une grande majorité de spécialistes, il en va tout autrement de ses nouvelles hypothèses sur l'Eurasiatique ou l'Amérinde. Elles sont d'abord l'objet de critiques très virulentes de la part des indo-européanistes qui continuent à penser que l'IE ne peut être relié à aucune autre famille, car le changement linguistique est si rapide qu'après environ 6000 ans toute trace de relations antérieures est effacée par l'incessante érosion phonétique et sémantique.

La classification des langues aborigènes d'Amérique est aussi à l'heure actuelle un des domaines les plus controversés. A l'appui de sa famille amérinde, Greenberg présente plus de 300 séries de mots apparentés, ainsi que plusieurs douzaines d'éléments grammaticaux dont le fameux système pronominal amérindien *na/ma* "je, moi / tu, toi", mais les américanistes orthodoxes considèrent ces exemples comme des emprunts dus à des relations de contact et de multilinguisme. De plus, on aurait trouvé le même système *n/m* en Papouasie-nouvelle-Guinée, en Australie, voire en Afrique.

De même, les hypothèses sur l'Eurasiatique (dont le système pronominal est basé sur *m* pour la première personne et *t* pour la deuxième personne), sur le Déné-caucasien et sur l'Austro-asiatique restent également spéculatives.

Malgré les polémiques et l'absence de certitudes, toutefois, la conception des "unificateurs" (l'école américaine de Greenberg et Ruhlen, ainsi que l'école russe de Dolgopolski et Starostin) compte aujourd'hui paradoxalement de plus en plus d'adeptes. D'aucuns suggèrent même d'aller encore plus loin dans les regroupements de familles linguistiques. Des rapports pourraient être établis entre l'Amérinde et l'Eurasiatique, comme le suggère Ruhlen (1992). L'idée que l'Austrique, le Déné-caucasien, voire l'Indo-pacifique pourraient être proches n'est pas non plus saugrenue et mériterait sans doute des recherches complémentaires. Et de remonter ainsi à une proto-langue unique.

A l'instar de tous les humains, dont on pense qu'ils ont des ancêtres communs (les fameux T-MRCA "the most recent common ancestors" que Underhill *et al.* [2000] datent de 140.000 à 40.000 années), toutes les langues pourraient ainsi avoir une origine commune. Elles seraient apparentées, à des degrés divers, les unes aux autres, et seraient issues d'une même proto-langue. Ruhlen va même jusqu'à identifier un certain nombre de mots qui se retrouvent dans toutes les macro-familles de langues pour désigner à-peu-près la même chose (exemples : *tik* "doigt, un" et *pal* "deux") et il conclut qu'il est peu vraisemblable qu'un tel phénomène de convergence soit dû à des emprunts.

Le plus grand défi lancé à la typologie linguistique reste évidemment l'organisation en sous-familles de cette famille linguistique unique. Pour l'instant, l'existence de macro-familles, et a fortiori d'une seule proto-langue, est loin d'être acceptée.

Une étude de la situation telle qu'elle se présente en Asie Orientale et en Asie du Sud-Est révèle bien la variété des hypothèses de regroupements possibles et la difficulté d'avoir des arguments indiscutables en faveur de l'une d'entre elles.

Langues d'Asie orientale et d'Asie du sud-est

Les grandes familles

On considère habituellement que cette vaste zone géographique regroupe cinq grandes familles linguistiques : le sino-tibétain, l'austronésien, l'austro-asiatique, le tai-kadai et le miao-yao (hmong-mjen).

- Le sino-tibétain (ST) est daté approximativement de 4500 ans avant J.-C. On le divise généralement en deux branches qui auraient divergé il y a 6000 ans, à une période à laquelle les archéologues identifient la formation de ce qu'ils appellent "la Chine initiale" (Wang 1996, Chang 1986) : les langues sinitiques et les langues tibéto-birmanes (TB). Le TB regroupe quelques 250 langues, dont le tibétain, pour lequel les premiers documents écrits datent du 7^{ème} siècle, le birman (dont les premiers textes attestés sont du 12^{ème} siècle)³, le lolo, le jingpo, etc. Le rapprochement entre le birman et le tibétain date du 18^{ème} siècle. Les langues sinitiques, quant à elles, se divisent en sept groupes : chinois du nord ou mandarin, xiang, gan, wu, min, keja (hakka), yue (cantonais)⁴. Les premières inscriptions chinoises, sur os et carapaces de tortue, remontent au 14^{ème} siècle avant J.-C.

Le ST reste une hypothèse très discutée. Le mot "sino-tibétain" lui-même, forgé par Jean Przyluski en 1924, a été introduit en anglais en 1931 (*Sino-tibetan*) pour remplacer la théorie de l'indo-chinois (*Indo-chinese*). Une hypothèse concurrente (van Driem 1997, 2001), entre autres, assume ainsi qu'il n'y a aucune raison de considérer les langues sinitiques comme ne faisant pas partie du TB, auquel cas le ST n'existe pas. Les langues sinitiques seraient simplement une branche du TB,

voire même une sous-sous-sous-branche, selon le schéma proposé ci-dessous par van Driem (2001) :

Tibeto-Burman

Western (Brahmaputran and other primary taxa)

Eastern

Northern (Sino-Bodic)

Northwestern (Bodic)

Bodish

Himalayan

Northeastern (Sinitic)

Southern

Deep Southern (Burmese, Karenic)

Lolo-Burmese

Karenic

Central (Qiangic, Xifan)

L'austronésien (AN) daterait approximativement de 5000 avant J.-C. Il s'agit d'une des plus anciennes familles identifiées, avant même l'indo-européen, qui regroupe plus de 1200 langues. On ne possède néanmoins pas de trace écrite avant 670. Blust (1997) considère que l'AN compte dix branches, dont neuf, pour 26 langues seulement, sont à Taiwan : ayatal, saisiat, bunun, tsou, rukai, paiwan, ami, puyuma et yami. La dixième branche rassemble plus de 1100 langues, de Madagascar à l'est de la Polynésie. Cf. également Diamond (2000).

Pour Sagart (2001b), il n'y aurait que six branches en AN, toutes représentées à Taiwan, et une seule d'entre elles serait aussi représentée ailleurs. Dans ce scénario, les langues malayo-polynésiennes ne seraient plus qu'une sous-sous-sous-sous-branche d'un groupe paiwan-polynésien.

La famille austro-asiatique (AA), identifiée vers 1880, compte près de 150 langues. Elle se subdivise en deux branches : les langues munda du nord de l'Inde et les autres langues môn-khmer, notamment au Vietnam et au Cambodge. Comme l'AN, l'AA daterait de 4000 à 5000 ans avant J.-C. Son foyer pourrait se situer au sud-ouest de l'Asie du Sud-Est, peut-être même en Inde orientale, mais certainement pas en Chine (Diffloth 2001).

Le tai-kadai (daïque) regroupe les langues tai et les langues kadai. La famille comprend le thai, bien sûr, dont les plus anciens documents écrits remontent à 1292, mais aussi le lao, le zhuang, le buyi et les langues kam-sui (kam, sui, mulao, maonan, etc.) du sud de la Chine. Au début du 19^{ème} siècle, les langues tai étaient dans un vaste ensemble dit "indo-chinois", qui comprenait aussi dans un premier temps les langues môn-khmer, et qui se réduisit, à la fin du siècle, à deux branches majeures, le tibéto-birman et le sino-tai. Ce n'est que vers le milieu du 20^{ème} siècle que le tai cessa d'être associé au chinois. Comme le suggère Comrie (2000), c'est sans doute l'absence de morphologie affixale dans ces langues qui a conduit à l'erreur de relier généalogiquement des langues comme le thai et d'autres comme le chinois.

La famille tai-kadai comprend le thai, bien sûr, mais aussi le lao et les langues kam-sui (zhuang, zhongjia, mulao, maonan, etc.)

La famille miao-yao (hmong-mjen), enfin, comprend une trentaine de langues qu'on répartit en deux groupes : celui des langues miao (hmong du Guizhou, du Yunnan, d'Indochine, hmn de l'ouest du Guizhou, qo xiong de l'ouest du Hunan) et celui des langues yao (mjen, mun, tsao min).

Les macro-familles

Divers regroupements entre les cinq familles détaillées ci-dessus ont été proposés au cours du siècle dernier. Presque toutes les combinaisons ont été essayées. Certaines propositions sont devenues aujourd'hui obsolètes. Il en est ainsi, par exemple, de la famille sino-tai défendue par Maspero.

Au cours des quinze dernières années, cependant, des indications d'ordre archéologique, génétique, anthropologique et linguistique ont été accumulées, qui inclinent à penser que plusieurs familles linguistiques d'Asie Orientale et d'Asie du Sud-Est peuvent être reliées.

Si on exclut l'hypothèse d'un Sino-indo-européen, encore très fragile (Pulleyblank 1996), quatre scénarios de quatre macro-familles sont aujourd'hui en discussion : Austro-Tai, Austrique, Sino-caucasien, Sino-austronésien (devenu récemment Sino-tibétain-austronésien).

En ce qui concerne l'Austro-Tai, Schlegel (1901) avait déjà noté des rapports entre les langues tai et les langues austronésiennes. Mais c'est Benedict qui a véritablement proposé, dès 1942, l'existence d'une macro-famille Austro-tai, comprenant deux branches : le tai-kadai et l'austronésien, auxquelles on a parfois ajouté le miao-yao. Cette hypothèse est aujourd'hui toujours défendue par Ostapirat (2001) qui a établi des correspondances régulières entre le lexique commun du tai-kadai et l'austronésien. Elle est critiquée, notamment par Thurgood (1994) qui soutient que le vocabulaire austronésien dans les langues tai a été emprunté. Sagart (2001b) pense que les éléments austronésiens indiscutables du kadai ont leur origine dans un vieil austronésien qu'il appelle "austronésien ancêtre du kadai" (AAK), lui-même une branche du proto-AN et un parent proche du proto-malayo-polynésien. Aucune date n'est donnée pour l'Austro-tai, mais on peut penser que la macro-famille pourrait remonter à 6000 à 7000 avant J.-C.

L'Austrique est une hypothèse émise aussi il y a longtemps par Schmidt (1905) qui regroupait les familles AA et AN, datant l'une et l'autre de 4000 à 5000 avant J.-C. Elle est aujourd'hui reprise par Blust (1996) et par Reid (1994) qui ont présenté bon nombre de rapprochements entre les deux familles, dans les domaines lexicaux, mais aussi morphologiques. D'autres y ajoutent le miao-yao et les langues tai-kadai. Diffloth (2001) pense que ces rapprochements sont loin d'être convaincants. Il situe par ailleurs le proto-AA très au sud du sous-continent de l'Asie du sud-est, voire en Inde même, ce qui rend l'hypothèse de l'Austrique assez invraisemblable.

Après avoir identifié un nombre important d'items lexicaux communs entre le ST, le caucasien du nord et le inénisséen (langue ket), Starostin propose au début des années 1980 l'existence d'une macro-famille, le Sino-caucasien (Starostin 1989). Plus tard, Bengston (1991) y ajoute le basque, le bouroushaski et le na-déné, pour en faire la macro-famille appelée par Ruhlen le Déné-caucasien.

Au début des années 1990, enfin, Sagart suggère que le chinois et l'austronésien soient apparentés (Sagart 1993). Il relève en effet des correspondances phonétiques régulières ainsi qu'un vocabulaire et une morphologie communs à ces langues, ce qui tend à prouver qu'elles sont généalogiquement reliées. Il affirme en conséquence l'existence d'une macro-famille, le Sino-austronésien, divisée en deux branches : les langues sinitiques et les langues austronésiennes. Le tibéto-birman est laissé de côté, ce qui sera l'objet de plusieurs critiques.

Il revient aujourd'hui sur cette hypothèse en reconnaissant une intégrité au sino-tibétain (avec ses deux branches, langues sinitiques et langues TB) et propose une

nouvelle macro-famille : le Sino-tibétain-austronésien (STAN) (Sagart 2001a). La raison essentielle est la suivante : la plupart du vocabulaire de base commun au chinois et à l'AN est aussi attesté en tibéto-birman. Les langues tai-kadai sont aussi incluses dans le STAN, qui daterait de 6500 environ avant J.-C., au moment de la domestication des céréales dans la vallée du Fleuve jaune.

Une macro-macro-famille : le PEA (proto Eastasian)

Les liens présumés existant entre les diverses familles linguistiques dans les hypothèses évoquées ci-dessus de macro-familles ont naturellement conduit à supposer une langue ancestrale commune pour toutes ces langues. Le pas a été franchi par Straosta (2001), qui appelle cette macro-macro-famille le Proto-Eastasian (PEA).

Le PEA aurait été parlé en Chine centrale, autour de la rivière Han et du Fleuve jaune, aux alentours de 10000 à 8000 avant J.-C. Il comprendrait seulement deux branches : le STAN (en reprenant la suggestion de Sagart (2001a) d'y inclure aussi le tai-kadai) et l'austro-asiatique. Nul besoin alors de reconnaître une macro-famille austrique.

Cette hypothèse d'une seule macro-macro-famille pour l'Asie Orientale et l'Asie du Sud-Est reste une pure spéculation, mais elle pourrait être soutenue par les études purement typologiques, voire de linguistique aréale, qui utilisent plus volontiers des critères syntaxiques que des critères d'ordre phonétique ou lexical. Bon nombre de propriétés syntaxiques ont en effet été identifiées, qui sont communes à la plupart des langues de cette aire très vaste (Matisoff, 1991, Chappell 2001, Enfield 2002, Chappell et Peyraube 2002) : développement de verbes modaux en marqueurs du désiratif, de verbes signifiant "obtenir, acquérir" en auxiliaires modaux du déontique et de l'épistémique, de verbes performatifs (du type "dire") en complémentiseurs et marqueurs du conditionnel ; grammaticalisation de verbes ayant le sens de "résider, habiter" en marqueurs aspectuels du progressif, de verbes signifiant "finir, achever" en marqueurs aspectuels de l'accompli, de verbes "donner" en marqueurs du datifs, du bénéfactif et du causatif, voire en marqueurs du passif ; formation de composés directionnels et résultatifs par concaténation verbale, à partir de constructions verbales en série. On pourrait aisément compléter cette liste. Ces traits grammaticaux, toutefois, servent à caractériser une zone, une aire linguistique. Ils ne sauraient être des critères pour une classification proprement généalogique des langues.

De fait, il existe bien plusieurs systèmes de classification des langues, qui aboutissent à des résultats différents.

Les différents systèmes de classification des langues

Comme tout ensemble d'objets, les langues peuvent être classées selon divers critères ou combinaison de critères. Ces critères classificatoires sont des propriétés ou des ensembles de propriétés tels que tous les éléments qui les possèdent appartiennent à un même ensemble alors que ceux qui ne les possèdent pas sont dans des ensembles différents, étant entendu que chaque élément appartient à un seul ensemble.

Quand on parle de classification des langues, on entend généralement une classification généalogique ou historique. C'est là le point de vue orthodoxe. Une telle classification opère avec des niveaux hiérarchiques. On la représente habituellement sous la forme d'arbres de familles. Le français, par exemple,

appartient au groupe des langues romanes, lui-même dépendant de la famille indo-européenne, que les "unificateurs" peuvent aussi rattacher à la macro-famille eurasiatique. Il y a cependant d'autres approches de classement des langues, d'autres systèmes de classification.

On peut distinguer trois approches fondamentales : l'approche généalogique qui conduit à construire des arbres (modèle du *Stammbaum*), l'approche aréale qui privilégie les contacts géographiques et entraîne à raisonner en termes d'extension et de recouvrements d'aires et l'approche dite typologique qui cherche les similarités et les ressemblances des structures linguistiques en mettant aussi au jour les divergences les plus importantes (Szulmajster-Celnikier 1998)

Approche généalogique

Les linguistes ont longtemps privilégié l'approche généalogique, dite aussi comparative parce qu'elle s'est imposée avec la grammaire historique et comparée des langues indo-européennes, sur un terrain où se sont enracinées les conceptions évolutionnistes de l'époque (Laks 2002, Philippson, dans ce volume, Greenberg 1992). Les familles sont organisées hiérarchiquement en langues mères, langues filles et langues sœurs. La méthode est fondée sur la statistique lexicale portant sur un vocabulaire de base, et sur l'identification de correspondances phonétiques régulières, étant entendu que le changement des sons est régulier, ce qui permet de rendre les reconstructions proposées plus convaincantes. "Pour les comparatistes, c'est en définitive la loi phonétique qui constitue la preuve interne de la validité de la reconstruction" (Laks 2002). Ce mode de classification est toujours considéré comme étant supérieur aux autres (Greenberg 2001).

L'approche généalogique part du mot, unité naturelle des langues. Elle reconstruit le son et le sens. Le problème est que les langues ne conservent pas leur stock lexical et morphologique originaux. Au cours de l'histoire des langues, les items lexicaux ou les affixes, même les plus basiques, disparaissent peu à peu et sont remplacés par des éléments nouveaux. Plusieurs linguistes pensent qu'après 10000 ans tout au plus, les mots d'une langue mère qu'on peut éventuellement retrouver dans des langues filles sont si peu nombreux, qu'il n'est pratiquement plus possible de les distinguer des ressemblances proprement accidentelles (Ringe 1999).

La méthode comparative met en œuvre des comparaisons bilatérales entre deux langues. A priori, des comparaisons multilatérales devraient pouvoir être menées simultanément, en partant des items les plus stables du lexique, par exemple les pronoms. Une telle approche est impossible, et à dire vrai, elle n'est pas nécessaire. Il suffit en effet de comparer entre elles des familles ou, mieux encore, des proto-langues ou des macro-familles (Greenberg 2001).

Approche aréale

La classification dite aréale privilégie les considérations d'ordre géographique. Les faits homologues observés au sein de la diversité des langues ne s'expliquent pas par une origine commune. Ils sont dus simplement au voisinage et aux relations continues qui s'ensuivent. Cette conception a entraîné des révisions importantes de certaines reconstructions. C'est ainsi que l'unité du groupe ouralo-altaïque (langues ouraliennes et langues altaïques) n'est plus reconnue. Il en est de même du chamito-sémitique (langues sémitiques et certaines langues d'Afrique saharienne et orientale, dont l'égyptien).

Il y a plus d'un siècle, Schmidt proposait déjà, comme alternative au *Stammbaum*, que les innovations dialectales se répandaient comme des ondes (*Wellen-Theorie*)⁵.

Les mélanges linguistiques pourraient être ainsi centraux pour la construction d'une typologie des langues adéquate. Baudouin de Courtenay, plus tard, défendait déjà avec force "le caractère mixte de toutes les langues". Jakobson (1938) parlait aussi de "théorie des affinités linguistiques". Si l'hybridation règne ainsi dans les langues, c'est assurément parce qu'elles ne sont pas des organismes vivants classables par des arbres de descendance, mais des systèmes sociaux, culturels et historiques qui s'interpénètrent et se mélangent (Laks 2002).

Cette méthodologie pourrait permettre de caractériser une vaste chaîne en Asie Orientale et Asie du Sud-Est regroupant le sino-tibétain, l'austronésien, le miao-yao, l'austro-asiatique et le tai-kadai (voir plus haut l'hypothèse du PEA (Proto-Eastasian)).

On définit plus précisément une aire linguistique par un ensemble de traits communs qui caractérisent les langues de cette aire et qui ont été introduits dans la langue par emprunt, après contact. Ces traits peuvent être à la fois aréaux et typologiques (voir *infra*), mais ils ne peuvent être en même temps généalogiques et aréaux, car cela impliquerait deux scénarios historiques mutuellement exclusifs. Comme les aires linguistiques les plus significatives sont le résultat historique de diffusions, les plus solides sont celles dont on a pu montrer que les traits qui les caractérisent ont été diffusés d'une zone à une autre et ne sont pas inhérents à une proto-langue ancestrale commune (Campbell, Kaufman & Smith-Stark 1986).

Un exemple classique d'aire linguistique est la zone des Balkans, où coexistent des langues généalogiquement différentes comme le roumain, l'albanais, le bulgare, le grec moderne, et qui forment pourtant une "confédération de langues" (*Sprachbund*) (Comrie 1989, p. 205). Un autre exemple est l'aire du bassin Içana-Vaupés, dans le nord-est de l'Amazonie, regroupant des langues de trois familles généalogiquement différentes (arawak du nord, tucano-est et maku) (Aikhenvald 1999).

Si on regarde sur une carte la distribution géographique de traits linguistiques, on se rend compte souvent que des langues proches géographiquement, qui partagent un trait linguistique majeur, en partagent beaucoup d'autres, même si elles appartiennent généalogiquement à des familles différentes.

Il reste qu'il est souvent difficile de distinguer, dans les ressemblances entre les langues, celles qui sont le résultat d'emprunts par contact de celles qui pourraient être explicables par une même source. En l'absence de données historiques, comme c'est souvent le cas, origine commune et contagion sont deux hypothèses envisageables.

Approche typologique

Il existe un troisième mode de classification : la méthode dite typologique. Les langues sont rangées en types, en se fondant sur les similarités des structures linguistiques, la plupart du temps syntactico-sémantiques. On peut ainsi distinguer des langues à distinction verbo-nominale (IE), à relative indistinction verbo-nominale (santali en Inde) ou mixtes (ST) ; des langues à construction ergative, marquant formellement l'agent (basque ou langues du Sud du Caucase), langues à construction accusative, marquant formellement le patient (français ou malais), langues mixtes (océaniques ou australiennes). Cf. Hagège (1982, pp. 39-40).

La classification ancienne en langues agglutinantes (constituées de radicaux et d'affixes identifiables, comme le turc), langues flexionnelles (avec des verbes conjugués, des noms déclinés, comme en russe), et langues isolantes (souvent

monosyllabiques et tonales, comme le chinois où les mots sont invariables) est une classification par types. Elaborée au 19^{ème} siècle, cette approche typologique a aujourd'hui beaucoup évolué. Elle a montré qu'en de nombreux endroits une grande diversité généalogique s'accompagnait de traits typologiques communs, rejoignant en cela des résultats obtenus par la méthode de classification aréale. De fait, familles de langues établies selon la méthode généalogique et types linguistiques coïncident rarement.

C'est une approche de ce type qu'utilise Nichols (1992) qui critique la méthode comparative standard, incapable, selon elle, de remonter au-delà de - 6000 années. Pour élaborer des hypothèses de regroupements plus intéressantes pouvant aller jusqu'à - 30000 années, elle propose d'utiliser, à partir d'un corpus de 174 langues, ce qu'elle appelle des "traits structurels stables" comme l'ordre des mots, les relations "tête-modifieur", la complexité morphologique ou les relations sujet-objet dans la proposition (ergativité, accusativité, etc.) (Nichols 1992, pp. 45 *sq.*, 1997).

Un des critères les plus utilisés aujourd'hui d'une classification typologique des langues est sans doute l'ordre des constituants dans la phrase. Des corrélations ont aussi été remarquées par Greenberg (1963b), qui permettent de définir plus précisément des types de langues. Ainsi, si une langue est SOV (sujet-objet-verbe), elle a des postpositions et le modifieur précède la tête dans des constructions endocentriques. Si elle est SVO, en revanche, elle a plutôt des prépositions et un ordre "tête + modifieur".

Les classifications typologiques les plus intéressantes et les plus productives se doivent d'établir des corrélations entre diverses typologies et suggérer qu'il puisse ainsi y avoir des liens causaux de nature universelle entre les propriétés considérées (existence d'"universaux implicationnels"). C'est pour cette raison que les travaux typologiques les plus récents ont été étroitement associés à la recherche d'universaux linguistiques, et que les typologies multi-dimensionnelles sont de plus en plus fréquentes.

Un des enseignements principaux de la typologie est sans doute le suivant : en dépit de la très grande diversité des langues, les différences restent contenues dans certaines limites, et les ressemblances de fond, fréquentes, désignent sans doute les capacités langagières de l'homo sapiens moderne. Elles restent cependant encore aussi peu interprétables que l'expansion de l'art paléolithique dans les mêmes formes et quasi simultanément en Tanzanie, Afrique de l'Est, Afrique du Sud, Australie, Inde et Amérique.

Les principaux domaines de divergence des langues étant supposés être, par ordre croissant, la syntaxe, la phonologie et le lexique, l'idée a été émise que les critères syntaxiques pouvaient être plus fiables que les critères phonologiques ou lexicaux pour la reconstruction de proto-langues. C'est sans doute aussi une illusion.

On a souvent considéré, par exemple, que l'ordre des mots d'une langue était un trait structural stable, imperméable à l'emprunt externe. Rien n'est plus faux (Peyraube 1986, 2002). On connaît de nombreux exemples de changements d'ordre des mots dus à des facteurs internes, mais aussi à des emprunts. Thomason et Kaufman (1988, p. 55) affirment même que l'ordre des mots est un des traits syntaxiques les plus aisément empruntables d'une langue à une autre. Ils citent les cas de SOV > SVO en finnois sous l'influence de l'indo-européen, en maa' sous l'influence des langues bantoues, les cas de VSO > SOV en akkadien sous l'influence du sumérien, les cas de SVO > SOV dans des langues austronésiennes de Papouasie-Nouvelle-Guinée sous l'influence des langues papoues. L'ordre des

mots du quetchua de Cochabamba-ville (en Bolivie) est de même en train de changer sous la pression de l'espagnol : le verbe, dans la plupart des phrases, n'est déjà plus en position finale (Hagège 2000, p. 101).

Langues et gènes

La génétique des populations, qui se propose, à partir de l'étude de la variabilité génétique de notre espèce, de reconstruire l'histoire des humains modernes depuis leur origine jusqu'à nos jours, est venue prêter main forte aux linguistes dans leur entreprise de classification des langues. Dès que les généticiens ont commencé, dans les années 1980, à pratiquer systématiquement l'analyse de l'ADN, les données se sont en effet considérablement accumulées, en même temps que se sont développées les techniques bio-chimiques et moléculaires, ainsi que les modèles théoriques d'évolution nécessaires à l'interprétation du matériel génétique dans un contexte historique.

Des corrélations entre distance génétique - notion centrale en génétique des populations - et distance linguistique ont été alors tentées. Et des correspondances étroites entre la classification génétique des populations et celle des macro-familles de langues, telle qu'elle est proposée par Greenberg et Ruhlen ont été trouvées (Greenberg *et al.* 1986 sur les populations américaines, Excoffier *et al.* 1987 sur l'Afrique sub-saharienne, Sokal *et al.* 1988 et Barbujani *et al.* 1990 sur l'Europe, Poloni *et al.* 1997 sur plusieurs populations d'Afrique et d'Europe, et surtout Cavalli-Sforza *et al.* 1988 qui ont construit un arbre de différenciations de quarante-deux populations humaines, issues de continents différents).

Il est vrai que les hypothèses linguistiques les plus volontiers retenues par les généticiens sont celles des macro-familles. Cela s'explique par le fait que les recherches génétiques sont le plus souvent réalisées sur de larges zones géographiques, à la fois pour des raisons d'échantillonnage et parce que les variations locales des fréquences géniques sont faibles. L'évolution génétique est aussi beaucoup plus lente que le changement linguistique et le parallélisme entre génétique et linguistique n'a de sens que s'il concerne des divergences très anciennes.

D'autres travaux ont ensuite infirmé l'existence de corrélations indiscutables entre classification génétique des populations et classification des langues. Cavalli-Sforza lui-même a reconnu ainsi que la corrélation entre distance génétique et distance linguistique est "extrêmement faible, sinon nulle" (Cavalli-Sforza *et al.* 1992, p. 5623). Il a aussi remarqué que les arbres génétiques et linguistiques des Mélanésiens ne correspondaient pas. En effet, les Mélanésiens sont groupés génétiquement avec les Micronésiens et les Polynésiens. La distance génétique avec les habitants de Nouvelle-Guinée est beaucoup plus grande. Linguistiquement, pourtant, les Mélanésiens et les habitants de Nouvelle-Guinée parlent des langues de la famille Indo-Pacifique alors que les Polynésiens et les Micronésiens parlent des langues austronésiennes (Cavalli-Sforza 2000).

On connaît aussi aujourd'hui deux cas, dans le Caucase, où les corrélations ne sont pas bonnes. Le premier concerne les Arméniens et les Azeris. Ils parlent des langues différentes (l'arménien est une langue indo-européenne et l'azeri une langue turque, donc altaïque) mais sont néanmoins très proches génétiquement. Le deuxième cas est inverse : les Tchétchènes et les Ingouches parlent des langues très voisines (appartenant à la branche des langues du Nord-Caucase) mais sont très différents du point de vue génétique (Nasidze *et al.* 2001).

De même, si les données des systèmes génétiques Rhésus et GM analysées par Dupanloup *et al.* (2002) s'accommodent très bien de l'hypothèse d'une macro-famille eurasiatique, les données Rhésus sont loin de soutenir, de près ou de loin, l'existence d'une macro-famille austrienne. Dans ce dernier cas, en effet, la proximité génétique des populations de langues austro-asiatique, tai-kadai et miao-yao, indiscutable, va de pair avec une extrême hétérogénéité génétique relevée pour les populations austronésiennes, surtout d'Océanie, sans doute en raison de leur isolement insulaire.

Mais qu'en est-il plus précisément de la situation en Asie Orientale et en Asie du Sud-Est ?

Langues et gènes en Asie Orientale et en Asie du Sud-Est

Un des premiers travaux importants menés par les généticiens chinois a étudié la distribution des allotypes de l'immoglobuline Gm et Km dans 74 populations de Chine Populaire⁶. Les articles de Zhao et Lee (1989), et surtout de Zhao *et al.* (1991) concluent clairement que les Han du Nord et les Han du Sud manifestent des différences génétiques marquées. La division passe le long d'une ligne correspondant à la latitude 30° Nord.

Une comparaison avec 33 autres populations hors de Chine fait apparaître que les Han du Nord appartiendraient volontiers à un groupe comprenant aussi les Althabasques (parlant des langues na-déné d'Amérique du Nord, i.e. de la macro-famille Déné-caucasienne), les Esquimaux, les Japonais, les Coréens et les Mongols (parlant des langues eskimo-aléoutes ou altaïques appartenant à la macro-famille Eurasiatique). Les Han du Sud, en revanche, seraient plus proches des Thaïlandais (langues tai-kadai), des Indonésiens et des Philippins (langues austronésiennes), enfin des Vietnamiens (langue austro-asiatique), bref autant de populations dont les langues sont subsumées dans la macro-famille Austrienne.

Chu *et al.* (1998), après avoir étudié le profil génétique de 28 populations chinoises, soutiennent fermement cette distinction Nord-Sud. Il en est de même de Su *et al.* (1999), dont les analyses ont porté sur le chromosome-Y. Ils notent que les populations du sud révèlent beaucoup plus de polymorphismes que celles du nord, qui partagent seulement avec elles un sous-ensemble limité d'haplotypes. L'unité des Han serait donc culturelle et non biologique. Elle pourrait être aussi non linguistique, les langues sinitiques étant rattachées, c'est selon, au Déné-caucasien quand il s'agit des langues parlées dans le Nord de la Chine ou à la macro-famille Austrienne pour le chinois du sud.

Une telle hypothèse, qui fait éclater l'ensemble sino-tibétain, mais aussi le groupe des langues sinitiques, serait presque conforme à la fois avec la théorie de Starostin sur le sino-caucasien (dans un ensemble Déné-caucasien) et avec celle de Starosta sur le PEA.

Ding *et al.* (2000), cependant, après avoir analysé trois marqueurs et un virus humain, affirment qu'une telle distinction nord-sud n'a pas de raison d'être. Ils notent ainsi que les populations zhuang du Sud de la Chine, qui parlent le zhuang (une langue tai-kadai) sont très proches génétiquement des Chinois du Nord, beaucoup plus proches qu'ils ne le sont des Vietnamiens (locuteurs de langues austro-asiatiques) ou des Malais (locuteurs de langues austronésiennes). Ils concluent que les langues observées entre Chinois septentrionaux et méridionaux sont dues simplement à des phénomènes culturels sans implication quelconque au niveau de la génétique.

Yao *et al.* (2002) pensent aussi qu'une simple division entre Chinois du Sud et Chinois du Nord est inappropriée pour rendre compte de la structure génétique des Han, autrement plus complexe⁷.

Conclusion

Comme le souligne Renfrew (2000), le débat sur l'existence ou non de macro-familles, sans parler de celle, encore plus hypothétique, d'une seule "langue-mère", n'est pas près d'être résolu. Tant qu'on n'aura pas montré que ces *macrophyllas* partagent plusieurs centaines de lexèmes primordiaux, il n'y aura pas d'évidences. Cependant, il est de plus en plus évident que la profondeur historique pour reconstruire des familles de langues ou des proto-langues ne se limite plus à - 6000. Dire au demeurant que le proto-indo-européen était parlé vers - 6000 est une pure supposition. On n'en sait tout simplement rien. Il aurait tout aussi bien pu être parlé il y a 15000 ans. De fait, la possibilité de trouver des correspondances entre la dispersion et l'organisation des langues et les processus démographiques qui ont suivi la fin du Pléistocène [- 10000 à - 8000] est de plus en plus grande. Les recherches les plus récentes et les plus prometteuses, quelles que soient les hypothèses contradictoires et l'absence de certitudes auxquelles elles ont abouti pour l'instant, ont ceci de particulier qu'elles sont devenues réellement interdisciplinaires. Les linguistes sont sortis de leur strict domaine de compétences et ont éprouvé le besoin de travailler avec des généticiens, mais aussi des archéologues, des paléoanthropologues, des paléodémographes, pour tenter de réaliser cette "nouvelle synthèse" que Renfrew appelle de ses vœux depuis le début des années 1990.

Références bibliographiques

- Aikhenvald A. (1999), "Areal diffusion and language contact in the Içana-Vaupés basin, North-West Amazonia", R. Dixon et A. Aikhenvald édés., *The Amazonian languages*. Cambridge : Cambridge University Press. 385-416.
- Barbujani G. et R. Sokal (1990), "Zones of sharp genetic change in Europe are also linguistic boundaries", *Proceedings of the National Academy of Science* 87 : 1816-1819.
- Benedict (1942), "Thai, Kadai, and Indonesian: A new realignment in Southeast Asia", *American Anthropologist* 44: 576-601.
- Bengston J.D. (1991), "Notes on Sino-caucasian", V. Shevoroshkin éd., *Dene-Sino-Caucasian languages*. Bochum : Studienverlag Brockmeyer.
- Blust R. (1996), "Beyond the Austronesian homeland : the Austric hypothesis and its implications for archaeology", W. Goodenough éd., *Prehistoric settlement of the Pacific*. Philadelphia : American Philosophical Society. 117-140.
- Blust R. (1997), "Subgrouping of the AN languages: consensus and controversies". Communication au 8^{ème} colloque international de linguistique austronésienne. Taipei.
- Campbell L., Kaufman T. et T. Smith-Stark (1986), "Meso-America as a linguistic area", *Language* 62-3 : 530-539.
- Cavalli-Sforza L.L., A. Piazza, P. Menozzi et J. Mountain (1988), "Reconstruction of human evolution : bringing together genetic, archaeological and linguistic data", *Proceedings of the National Academy of Sciences* 85 : 6002-6006.
- Cavalli-Sforza L.L., E. Minch et J. Mountain (1992), "Coevolution of genes and languages revisited", *Proceedings of the National Academy of Sciences* 89 : 5620-5624.
- Cavalli-Sforza L.L. (2000), *Genes, Peoples, and Languages*. Farrar, Strauss & Giroux.
- Chafe W.L. (1962), "Estimates regarding the present speakers of North American Indian Languages", *International Journal of American Linguistics* 28 : 162-171.
- Chang K.C. (1986), *The Archaeology of Ancient China*. New Haven : Yale University Press.
- Chappell H. (2001), "Language contact and areal diffusion in Sinitic languages", A. Aikhenvald & R. Dixon édés., *Areal diffusion and genetic inheritance*. Oxford : Oxford University Press. 328-357.

Chappell H. et A. Peyraube (2002), "A Diachronic exploration of analytic causatives in Early Southern Min". Communication au Colloque de linguistique chinoise de Nanchang (Jiangxi, Chine).

Chu J.Y., W. Huang, S. Kuang, J. Wang, J. Xu, Z. Chu, Z. Yang, K. Lin, P. Li, M. Wu, Z. Geng, C. Tan, R. Du et L. Jin (1998), "Genetic relationship of populations in China", *Proceedings of the National Academy of Sciences* 95 : 11763-11768.

Comrie B. (1989), *Language universals and linguistic typology*. Oxford : Basic Blackwell. 2nd edition.

Comrie B. (2000), "Is there a single time depth cut-off point in historical linguistics?", C. Renfrew, A. McMahon et L. Trask éd., *Time depth in historical linguistics*. Cambridge: McDonald Institute for Archaeological Research. 33-43.

Contini M. (2000), "Vers une nouvelle linguistique historique : l'ouvrage de Mario Alinei, *Origini delle lingue d'Europa*", *Dialectologia et Geolinguistica* 2000-8 : 13-35.

Diamond J.M. (2000), "Taiwan's gift to the world", *Nature* 403 : 709-710.

Diffloth G. (2001), "On the high cost of Austric". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.

Ding Y.C., S. Wooding, H. Harpending, H. Chi, H. Li, Y. Fu, J. Pang, Y. Yao, J. Yu, R. Moyzis et Y. Zhang (2000), "Population structure and history in East Asia", *Proceedings of the National Academy of Sciences* 97 : 14003-14006.

Dolgopolsky A. (1998), *The Nostratic macrofamily and linguistic paleontology*. Cambridge: McDonald Institute for Archaeological Research.

Dupanloup I., E. Poloni, S. Schneider, L. Excoffier et A. Langaney (2002), "Génétique, linguistique et histoire des peuplements humains", *Langages* 146 : 80-90.

Enfield N. (2002), *Linguistic epidemiology : Semantics and grammar of language contact in Mainland Southeast Asia*. London : Routledge.

Excoffier L, B. Pelligrini, A. Sanchez-Mazas, C. Simon et A. Langaney (1987), "Genetic and history of Sub-Saharan Africa", *Yearbook of Physical Anthropology* 30 : 151-194.

Greenberg J.H. (1963a), *Languages of Africa*. Bloomington : Indiana Research Center in Anthropology.

Greenberg J.H. (1963b), "Some universals of grammar with particular reference to the order of meaningful elements", J.H. Greenberg éd., *Universals of Language*. Cambridge : MIT Press.

Greenberg J.H. (1987), *Languages in Americas*. Stanford : Stanford University Press.

Greenberg J.H. (1992), "Preliminaries to a systematic comparison between biological and linguistic evolution", J.A. Hawkins et M. Gell-Mann édés., *The Evolution of human languages*. Redwood City (CA) : Addison-Wesley Publishing Company. 139-158.

Greenberg J.H. (2001), "The methods and purposes of linguistic genetic classification", *Language and Linguistics* (Taiwan) 2-2 : 111-135.

Greenberg J.H. (à paraître), *Indo-European and its closest relatives : the Eurasiatic language family*. Stanford : Stanford University Press.

Greenberg J.H., C. Turner, et S. Zegura (1986), "The settlement of the Americas : A comparison of the linguistic, dental, and genetic evidence", *Current Anthropology* 27 : 477-497.

Hagège C. (1982), *La structure des langues*. Paris : PUF.

Hagège C. (2000), *Halte à la mort des langues*. Paris : Odile Jacob.

Jakobson R. (1938), "Sur la théorie des affinités phonologiques entre les langues", *Actes du 4^{ème} congrès international des linguistes*. Copenhague. Repris dans Troubetzkoy N., *Principes de phonologie*. Paris : Klincksieck, 1947 : 351-365.

Laks B. (2002), "Le comparatisme : de la généalogie à la génétique", *Langages* 146 : 19-45.

Matisoff J. (1991), "Areal and universal dimensions of grammaticalization in Lahu", E. Traugott et B. Heine édés., *Approaches to grammaticalization*. Amsterdam : John Benjamins. 383-453.

Nasidze I., G. Risch, M. Robichaux, S. Sherry, M. Batzer et M. Stoneking (2001), "Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus", *European Journal of Human Genetics* 9 : 267-272.

Nichols J. (1992), *Linguistic diversity in space and time*. Chicago : University of Chicago Press.

Nichols J. (1997), "The epicentre of the Indo-European linguistics spread", R. Blench R. et M. Spriggs édés., *Archaeology and language I. Theoretical and methodological orientations*. London : Routledge. 122-148.

Ostapirat W. (2001), "Kra-dai and Austronesian : notes on some phonological correspondences". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.

Pedersen H. (1903), "Türkische Lautgesetze", *Zeitschrift der Deutschen Morgenländischen Gessellschaft* 57 : 535-561.

- Peyraube A. (1986), "Recent issues in Chinese historical syntax", J. Huang et A. Li éd.s., *New Horizons in Chinese Linguistics*. Dordrecht : Kluwer. 161-214.
- Peyraube A. (2002), "L'évolution des structures grammaticales", *Langages* 146 : 46-58.
- Philippson G. (dans ce volume), "La méthode comparative".
- Poloni E.S., O. Semino, G. Passarino, A. Santachiara-Benerecetti, I. Dupanloup, A. Langaney et L. Excoffier (1997), "Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondance with linguistics", *American Journal of Human Genetics* 61 : 1015-1035.
- Pulleyblank E.G. (1996), "Early contacts between Indo-Europeans and Chinese", *International Review of Chinese Linguistics* 1-1 : 1-24.
- Reid L. (1994), "Morphological evidence for Austric", *Oceanic linguistics* 33-2 : 323-344.
- Renfrew C. (2000), "At the edge of knowability : towards a prehistory of languages", *Cambridge Archaeological Journal* 10-1 : 7-34.
- Ringe D. (1999), "Language classification: Scientific and unscientific methods", B. Sykes éd., *The Human Inheritance : Genes, Language and Evolution*. Oxford : Oxford University Press. 45-74.
- Ruhlen M. (1992), "An overview of genetic classification", J.A. Hawkins et M. Gell-Mann éd.s., *The Evolution of human Languages*. Redwood City (CA): Addison-Wesley Publishing Company. 159-189.
- Ruhlen M. (1994), *On the origins of languages : Studies in linguistic taxonomy*. Stanford: Stanford University Press.
- Ruhlen M. (1997), *L'origine des langues*. Paris : Belin.
- Sagart L. (1993), "Chinese and Austronesian : evidence for a genetic relationship", *Journal of Chinese Linguistics* 21-1 : 1-62.
- Sagart L. (2001a), "The Evidence for Sino-Austronesian". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.
- Sagart L. (2001b), "Comment : Malayo-polynesian features in the AN-related vocabulary in Kadai". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.
- Schlegel G. (1901), "Review of Frankfurter's Siamese Grammar", *T'oung Pao* 2 : 76-87.

- Schmidt W. (1905), *Grundzüge einer Lautlehre der Mon-khmer Sprachen*.
- Sériot P. (1999), *Structure et totalité*. Paris : PUF.
- Sokal R.R., N. Oden et B. Thomson (1988), "Genetic changes across language boundaries in Europe", *American Journal of Physical Anthropology* 76: 337-361.
- Starosta S. (2001), "PEA: A scenario for the origin and the dispersal of the languages of East and Southeast Asia and the Pacific". Communication au symposium sur les perspectives d'une phylogénie des langues d'Asie Orientale. Périgueux.
- Starostin S. (1989), "Nostratic and Sino-Caucasian", V. Shevoroshkin ed., *Explorations in language macrofamilies*. Bochum: Studienverlag Dr. Brockmeyer. 42-66.
- Su B., J. Xiao, P. Underhill, R. Deka, W. Zhang, J. Akey, W. Huang, D. Shen, D. Lu, J. Luo, J. Chu, J. Tan, P. Shen, R. Davis, L. Cavalli-Sforza, R. Chakraborty, M. Xiong, R. Du, P. Oefner, Z. Chen et L. Jin (1999), "Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age", *American Journal of Human Genetics* 65 : 1718-1724.
- Szulmajster-Celnikier A. (1998), "Eloge de la prudence méthodologique", *La Recherche* 306 : 76-81.
- Thomason S.G & T. Kaufman (1988), *Language contact, creolization, and genetic linguistics*. Berkeley : University of California Press.
- Thurgood G. (1994), "Tai-kadai and Austronesian : the nature of the historical relationship", *Oceanic linguistics* 33-2 : 354-368.
- Trombetti A. (1905), *L'Unità d'origine del Linguaggio*. Bologne.
- Underhill P.A., P. Shen, A. Lin, L. Jin, G. Passarino, W. Yang, E. Kauffman, B. Bonnè-Tamir, J. Bertrandpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J. Kidd, Q. Mehdi, M. Seielstad, R. Wells, A. Piazza, R. Davis, M. Feldman, L. Cavalli-Sforza et P. Oefner (2000), "Y-chromosome sequence variation and the history of human populations", *Nature Genetics* 26 : 358-361.
- Van Driem G. (1997), "Sino-Bodic", *Bulletin of the School of Oriental and African Studies* 60-3 : 455-488.
- Van Driem G. (2001), *Languages of the Himalayas*. Leiden : Brill.
- Wang W.S.Y. (1996), "Genes, Dates, and the Writing System", *International Review of Chinese Linguistics* 1-1 : 45-50.
- Yao Y.G., Q. Kong, H-J Bandelt, T. Kivisild et Y. Zhang (2002), "Phylogeographic differentiation of mitochondrial DNA in han Chinese", *American Journal of Human Genetics* 70, à paraître.

Zhao T.M. et T. Lee (1989), "Gm and Km allotypes in 74 Chinese populations : a hypothesis of the origin of the Chinese nation", *Human Genetics* 83 : 101-110

Zhao T.M., G. Zhang, Y. Zhu, S. Zheng, W. Gu, Q. Chen, X. Zhang et D. Liu (1991), "Etude d'allotypes d'immoglobuline dans la population chinoise : hypothèse sur l'origine de la nation chinoise", *Acta Genetica Sinica* 18-2 : 97-108 (en chinois)

Notes

¹ Sur ces 5000 à 6000 langues, 600 seulement sont parlées par plus de 100.000 locuteurs, 500 le sont par moins de cent personnes (Hagège 2000). Enfin, 90 % d'entre elles sont utilisées par moins de 5 % environ de la population mondiale.

² La plupart des partisans du nostratique incluent aussi le kartvélien, le dravidien et l'afro-asiatique dans le nostratique, daté de 12000 à 9000 ans avant J.-C. Cf. Dolgopolski (1998).

³ On dispose néanmoins dans un opuscule des Han (ca. 2^{ème} siècle) d'un lexique de 155 mots d'une langue appelée "bailangge", manifestement tibéto-birmane : 110 de ces mots sont considérés comme du birman ancien et 40 d'entre eux comme du tibétain ancien.

⁴ Chappell (2001) parle de dix groupes de langues sinitiques. Aux sept groupes ci-dessus, elle rajoute les langues jin, hui et pinghua.

⁵ "Pour [Schmidt], les langues indo-européennes ne doivent pas être représentées comme des branches qui se séparent d'un tronc, mais comme une chaîne faite d'anneaux et n'ayant ni début, ni fin, ni centre, ni périphérie. Contrairement à ce que la théorie de l'arbre généalogique permettait de prévoir, des innovations apparues dans une langue séparée, distincte, peuvent s'étendre, se diffuser à d'autres langues voisines spatialement, comme des ondes qui se propagent à la surface de l'eau. Chaque onde en gagnant de proche en proche par un mouvement insensible, ne" comporte aucune limite". (Sériot 1999, p. 126).

⁶ La Chine est composée de 93,3 % de Chinois Han et de 6,7 % de nationalités non-Han, occupant toutefois à elles seules 55 à 60 % du territoire du pays.

⁷ Je remercie Lluís Quintana-Murci de m'avoir fourni un exemplaire des épreuves de l'étude de Yao *et al.*, non encore publiée au moment de la rédaction du présent article.