



L'origine des langues



<https://www.axl.cefap.ulaval.ca/monde/origine-langues.htm>

1 Monogénèse et polygénèse

La question de l'origine des langues a toujours suscité de nombreuses hypothèses et mis à contribution les travaux tant des anthropologues, que des archéologues, des généticiens, des linguistes, etc. En 1865, la Société de linguistique de Paris avait informé ses membres dans ses règlements qu'elle ne recevrait «aucune communication concernant [...] l'origine du langage». Mais la question a continué néanmoins à hanter les linguistes et la recherche d'une langue mère unique s'est poursuivie, si tant est qu'une telle langue ait existé. Dans *L'Homme de paroles* (Fayard, 1996), le linguiste français Claude Hagège réfute le mythe d'une langue commune unique :

Contrairement à l'idée courante, il est très probable que l'immense diversité des idiomes aujourd'hui attestés ne se ramène pas à une langue originelle unique pour toute l'humanité. S'il y a unicité, c'est celle de la faculté de langage propre aux hominiens et non celle de la langue elle-même. À l'origine, donc, une seule espèce (monogénéisme de la lignée), mais non un seul idiome (polygénisme des langues).

Néanmoins, l'idée d'une langue mère relève d'un fantôme ancien. Dès le Moyen Âge, on croyait à l'existence d'une langue originelle de l'humanité, jusqu'à ce que la colère de Dieu intervienne après l'épisode de la tour de Babel. Pendant longtemps, on a cru que l'hébreu était la langue d'Adam et d'Ève, d'autres ont pensé au latin ou au grec. Pour leur part, les musulmans ont toujours cru que la première langue de l'humanité était l'arabe. À partir du XIX^e siècle, certains linguistes ont persisté dans ce type de recherche; ils ont été suivis par des spécialistes de la génétique des populations. L'un des livres les plus connus sur le sujet fut celui de l'Américain Merritt Ruhlen (né en 1944) dans *L'origine des langues* (1994, mais 1997 en français). Ses travaux proposant une origine commune (la monogénèse) ont alimenté une controverse vieille de plusieurs siècles. Pour établir des ressemblances entre toutes les langues du monde, la méthode de Ruhlen consiste à procéder à des comparaisons entre des lexiques de référence (en l'occurrence: 27 formes orthographiques associées aux formes phonétiques) pour un grand nombre de langues choisies parmi des familles communément acceptées. Il s'agit du système de «comparaison multilatérale» proposée auparavant par le linguiste américain Joseph Greenberg (1915-2001).



Quoi qu'il en soit, **Merritt Ruhlen** a avancé la thèse d'une proto-langue mère originelle et commune à toutes les superfamilles, qui aurait vécu vers 50 000 ans avant notre ère. Selon lui, le premier mot prononcé par l'homme serait la monosyllabe *tik* («doigt») ou *aq'wa* («eau»), appartenant à 32 familles de langues et proto-langues reconnues par la majorité des linguistes. Cela étant dit, les critiques portant sur la méthodologie de Ruhlen sont innombrables. Non seulement on peut se demander si les ressemblances relevées par Ruhlen sont dues ou non au hasard, mais on met en doute la capacité des sons humains à se maintenir sur des dizaines de milliers d'années. Malgré tout, nombreux sont ceux qui reconnaissent au moins à Merritt Ruhlen le mérite d'avoir raison sur le fond: toutes les langues pourraient avoir une source unique, sauf que nous n'en savons strictement rien. L'origine des langues reste toujours une énigme pour la science.

Cependant, si le moment de l'émergence du langage demeure encore une énigme pour la science et divise les linguistes, il est généralement admis que l'aptitude au langage se soit inscrite il y a environ 2,2 millions d'années dans le code génétique de l'*Homo habilis*, dont la capacité à fabriquer des outils témoigne déjà d'une grande complexité de l'organisation neurologique. On croit que cette aptitude n'aurait été utilisée que bien plus tard par l'*Homo erectus*, sinon par l'*Homo sapiens sapiens*, selon les plus prudents. Les langues, dans leur sens moderne, ne seraient apparues qu'entre 80 000 à 60 000 ans avant notre ère, en Afrique de l'Est ou au Proche-Orient, alors que nos ancêtres, les *Homo sapiens sapiens*, n'étaient plus que quelques milliers d'individus. À supposer qu'ils aient pu parler, on peut se demander s'ils parlaient une langue commune — théorie de la monogénèse — auquel cas les quelque 6000 langues actuelles descendraient de cette langue parlée il y a 60 000 à 80 000 ans. On peut aussi imaginer que des langues existaient bien avant cette date et que les langues ne se soient développées qu'après la dispersion des différents groupes d'*Homo sapiens* (théorie de la polygénèse). Dans l'état actuel des choses, les outils de la science et de la linguistique comparée ne nous permettent pas d'en savoir davantage.

Pour sa part, le linguiste américain **Noam Chomsky** croit qu'il est possible qu'il y ait eu une langue d'origine unique, mais nous n'en savons strictement rien :

We don't come from Adam and Eve. Get your facts right. The story of Adam and Eve is completely false. Get out to the world and teach yourself some reality. As for the origin of languages, it is possible that languages have single origin. But, we don't have clear evidence yet. **[Nous ne venons pas d'Adam et Ève. Vérifiez vos sources. L'histoire d'Adam et Ève est complètement fautive. Sortez de votre monde et renseignez-vous sur une certaine réalité. Quant à l'origine des langues, il est possible que les langues aient une origine unique. Mais nous n'avons encore aucune preuve évidente.]**

La théorie néo-darwinienne de l'évolution plaide en faveur du polygénisme, c'est-à-dire que plusieurs couples humains seraient à l'origine de l'humanité. C'est au sein d'une espèce que prennent place les mutations génétiques, lesquelles séparent les espèces entre elles. Toutefois, cette théorie recèle encore beaucoup de zones d'ombre. On ne se surprendra pas que, dans ces conditions, la question sur l'origine des langues ne soit pas encore résolue.

Dernière mise à jour: 12 janvier, 2024





L'origine des langues



<https://www.axl.cefan.ulaval.ca/monde/origine-langues.htm>

2 Les méthodes classificatoires des langues

Il existe deux grands principes classificatoires des langues. Le premier s'intéresse aux **classements typologiques**; le second aux **classements génétiques**. La classification typologique des langues a pour but leur description et leur regroupement en fonction de certaines caractéristiques communes de leurs structures, sans rechercher nécessairement l'établissement de généalogies ou de familles de langues. La classification génétique s'intéresse plutôt aux familles de langues, c'est-à-dire à un ensemble de langues effectivement parentes, qui descendent d'une langue présumée commune ou originelle.

2.1 La méthode typologique

Dans les classements typologiques, les langues peuvent être caractérisées selon divers traits linguistiques. Par exemple, on peut classer les langues en fonction de critères phonétiques ou phonologiques, morphologiques ou syntaxiques.

- Les critères phonétiques ou phonologiques

C'est ainsi que l'on pourra distinguer les langues en fonction de leur système vocalique: les langues à trois voyelles ([i], [u], [a]), les langues à double articulation antérieure (voyelles non arrondies et arrondies), les langues à double articulation postérieures (voyelles non arrondies et arrondies), les langues à double durée vocalique, etc.

Du côté des consonnes, certains linguistes ont tenté de classer les langues en fonction des modes d'articulation: les langues à consonnes occlusives limitées (une seule), à consonnes fricatives limitées (seulement le [t]), les langues à consonnes pré-nasalisées ou post-nasalisées, les langues à clics, les langues à deux modes articulatoires, etc. D'autres distinguent les langues «à tons» (comme le chinois, le vietnamien, le birman), les langues «à accent tonique fixe» (tchèque, finnois, hongrois), les langues à accent tonique «à valeur phonologique» (russe), etc. Comme on le constate, il est possible d'en arriver à de nombreux types de classements. Le problème, c'est d'obtenir l'accord des spécialistes sur la question.

- Les critères morphologiques

Parmi les systèmes de classement typologique, les critères d'ordre morphologique semblent les plus connus. On distingue ainsi trois types principaux de langues: les langues **isolantes**, les langues **agglutinantes** et les langues **flexionnelles**.

a) Langue isolante

Il est admis de considérer une langue comme isolante lorsque les mots sont ou tendent à être invariables. En fait, une langue est isolante lorsque chacun des morphèmes est identifié à des mots graphiques isolables. Cela signifie que les marques du genre et du nombre, par exemple, constituent des morphèmes distincts et séparés du lexème, parce que chacun des mots correspond à un radical unique. Les langues isolantes les plus connues sont le chinois, le cantonais, le vietnamien, le laotien et le cambodgien. Voici un exemple en chinois:

ta	chi	fan	le	ta	chi	le	fan
il/elle	a mangé	repas/nourriture	"passé"	il	mange	passé	repas

= *Il/elle a pris son repas.*

b) Langue agglutinante

Dans une langue agglutinante, au contraire, on juxtapose au radical une série de morphèmes distincts servant à exprimer les rapports grammaticaux. Dans ce type de langue, chacun des affixes (préfixes, infixes ou suffixes) est clairement analysable et identifie précisément une fonction grammaticale ou syntaxique. En voici quelques exemples en turc, en quechua, en swahili et en créole haïtien.

En turc	En quechua	En swahili	En créole haïtien
ev = maison evim = ma maison evlerim = mes maisons evimden = de ma maison evlerimden = de mes maisons	wasi = (la) maison wasikuna = (les) maisons wasip = dans la maison wasikunap = dans les maisons wasiykikunap = dans tes maisons	penda = aimer anapenda = il aime atapenda = il aimera amependa = il a aimé atanipenda = il m'aimera amakupenda = il t'a aimé utanipenda = tu m'aimeras	li mangé = il mange (présentement) li ape mangé = il mange (intemporel) li te mangé = il a mangé li tap mangé = il mangeait li va mangé = il mangera li tava mangé = il aurait mangé li ta mangé = il mangerait

c) Langue flexionnelle

Enfin, dans une langue flexionnelle, les radicaux sont pourvus d'affixes grammaticaux variables et exprimant plus ou moins à la fois, par exemple, le genre, le nombre et le cas, ou la personne, le temps, le mode, la voix, etc. La plupart des langues européennes sont des langues considérées comme flexionnelles. Ainsi, en latin, la série *bonus dominus, boni domini, bonos dominos* oppose des morphèmes identifiant à la fois le nominatif (sujet) masculin singulier (*bonus dominus*), ou le génitif masculin singulier (*boni domini*), ou le nominatif masculin pluriel (*boni domini*), ou encore l'accusatif

masculin pluriel (*bonos dominos*). Référons-nous encore une fois au **système russe** qui oppose des terminaisons identifiant à la fois le cas, le genre masculin, féminin ou neutre, ainsi que le nombre.

MASCULIN: <i>dom</i> («maison») singulier - pluriel	FÉMININ: <i>ulica</i> («rue») singulier - pluriel	NEUTRE: <i>tchuvstvo</i> («sensation») singulier - pluriel
Nominatif : <i>dom domi</i> Génitif : <i>doma domov</i> Accusatif : <i>dom domi</i> Datif : <i>domu domam</i> Locatif : <i>dome domax</i> Instrumental : <i>domom domami</i>	Nominatif : <i>ulica ulicci-</i> Génitif : <i>ulici- ulic</i> Accusatif : <i>ulicu ulicci-</i> Datif : <i>ulice ulicam</i> Locatif : <i>ulice ulicax</i> Instrumental : <i>ulicoy ulicami</i>	Nominatif : <i>tchuvstvo tchuvstva</i> Génitif : <i>tchuvstva tchuvstv</i> Accusatif : <i>tchuvstvo tchuvstva</i> Datif : <i>tchuvstvu tchuvstvam</i> Locatif : <i>tchuvstve tchuvstvax</i> Instrumental : <i>tchuvstvom tchuvstvami</i>

Il arrive parfois que la variante flexionnelle soit interne; on parle en ce cas d'*infixe*. On aura, par exemple, en **anglais** *I drink, I drank, I have drunk* (je bois, je buvais, j'ai bu), et en **allemand** *Ich spreche, Ich sprach, Ich habe gesprochen* (je parle, je parlais, j'ai parlé).

Ces distinctions ne sauraient être considérées comme absolues et il conviendrait de parler en termes de degré. Le français est parfois de type flexionnel (cheval/chevaux), parfois de type isolant (je suis/tu es), parfois de type agglutinant (épais/épaisse). Un syntagme comme *porte-manteau* est isolant, alors qu'une opposition du genre *pomme/pommier* est flexionnelle. De même pour les cas suivants:

Français Espagnol Portugais Italien	<i>pomme</i> → <i>pommier</i> <i>manzana</i> → <i>manzano</i> <i>maça</i> → <i>macieira</i> <i>mela</i> → <i>melo</i>	Type flexionnel
Anglais Allemand Néerlandais Danois	<i>apple</i> → <i>apple tree</i> <i>Apfel</i> → <i>Apfelbaum</i> <i>appel</i> → <i>appelboom</i> <i>aeble</i> → <i>aebletrae</i>	Type isolant

Il importe donc de définir pour chacune des langues le caractère dominant, car une langue peut être plutôt flexionnelle et présenter, par exemple, des traits isolants et agglutinants.

- Les critères syntaxiques

Un autre critère pour classer les langues du monde est de recourir à l'ordre des mots dans la phrase. Autrement dit, c'est le critère syntaxique qui sert alors à distinguer les langues. Pour ce faire, on compare l'ordre du sujet, du verbe et du complément dans la phrase. Ainsi, en français, l'ordre syntaxique le plus courant est l'ordre sujet + verbe + complément (SVC). Voyons ce qu'il en est dans

d'autres langues, notamment en turc, en gallois, en malgache et en hixkaryana (langue amérindienne de l'Amazonie):

En français: [SVC] = Cet homme construit une maison.

Turc [SCV]	Gallois [VSC]	Malgache [VCS]	Hixkaryana [CVS]
Hasan Ėküz-ü ald-i.	Lladdodd y ddraig y dyn.	Nahita ny mpianatra ny vehivavy.	Toto yahosiye kamara.
Hasan boeuf (acc.) a acheté	tua le dragon l'homme	a vu l'élève la femme	homme saisit jaguar
«Hasan a acheté le boeuf.»	«L'homme tua le dragon.»	«La femme a vu l'élève.»	«Le jaguar saisit l'homme.»

Ces systèmes de classification typologique présentent certainement un intérêt, mais ils ont tous comme défaut principal de ne pas être très rigoureux. Pour cette raison, beaucoup de linguistes préfèrent classer les langues par famille, c'est-à-dire la méthode génétique.

2.2 La méthode génétique



La méthode génétique provient d'une conception biologique de la langue qu'avait adoptée le linguiste et philologue allemand **Franz Bopp** (1791-1867) au XIX^e siècle. Celui-ci s'était représenté les langues comme des êtres humains dont on pouvait suivre la naissance, la vie et la mort. Selon cette même conception, les langues avaient des «parents»; en ce sens, on parle de «langue mère», de «langues sœurs», de «langues cousines», etc. C'est dans cet esprit que le mot *génétique* a été appliqué à la linguistique. Aujourd'hui, ce mot est utilisé de plus en plus dans le sens de «historique»: lorsqu'on recherche des états de langue anciens, il est légitime de penser en termes d'affiliation et de parenté linguistiques.

- Liens de parenté linguistique

En analysant des milliers de langues parlées dans le monde, les linguistes ont pu établir certains liens de parenté plus ou moins étroits entre des parlers dont plusieurs peuvent représenter des évolutions différentes d'un même prototype (du grec *protos*: «premier», «primitif»). Généralement, on réserve l'expression **«famille linguistique»** à l'ensemble formé de toutes les langues de même origine (p. ex., la famille indo-européenne, la famille sémitique).

Ces familles comprennent des sous-ensembles appelés **«sous-familles»** ou **«branches»** (p. ex., la branche romane, la branche germanique, la branche slave, etc.). Ces branches sont elles-mêmes constituées de certaines langues plus étroitement apparentées entre elles qu'avec d'autres. Ainsi, les langues de la branche romane (français, espagnol, italien, espagnol, etc.) diffèrent de celles de la branche germanique (anglais, allemand, néerlandais, danois, etc.) et slave (russe, polonais, tchèque, slovène, etc.), mais elles appartiennent toutes à la même famille: la famille indo-européenne. Cette famille est ainsi appelée parce qu'elle regroupe un grand nombre de langues en usage depuis l'Inde (en passant par le Pakistan, l'Iran, l'Iraq, la Syrie et la Russie) jusqu'à l'Europe de l'Ouest (du Portugal à Moscou en passant par l'Islande et la Grèce).

On utilise aussi le terme de **«groupe»** (p. ex, les langues du groupe andino-équatorial de l'Amérique du Sud). Il s'applique indifféremment à un ensemble de familles, à une famille, à un ensemble de langues

d'une branche. L'utilisation de ce terme sous-entend que le classement n'est pas encore fixé ou n'est pas fixé de façon certaine.

- Des hypothèses

Il ne faudrait pas croire que l'établissement de liens de parenté entre les langues repose toujours sur une langue originelle véritable. Dans certains cas, il s'agit d'hypothèses que l'on formule d'après des analyses comparatives et historiques, afin de constituer des ensembles de langues. Les linguistes ont reconstitué des langues originelles, des protolangues, qui n'ont jamais été attestées et qui, pour cette raison, demeurent des langues purement hypothétiques. C'est le cas de l'**indo-européen**, reconstruit par les linguistes, car, étant donné qu'aucun document écrit ne peut confirmer son authenticité, on ne peut que supposer l'existence de cette langue.

Le seul fait dont on est sûr, c'est que, entre un certain nombre de langues diverses, entre un ensemble de traits communs remarquables, il existe une parenté indiscutable. On ignore comment était parlé l'indo-européen primitif, mais on connaît les langues qui en sont issues et ce qu'elles sont devenues en se différenciant de plus en plus avec le temps: le sanskrit en Inde, le vieux-perse en Iran et, en Europe, le grec, le latin, le celtique, le germanique, le slave, etc.

En général, la communauté scientifique a établi un consensus pour admettre l'existence d'environ 300 familles qui remonteraient au début de notre ère. Quant aux **macro-familles** (les superfamilles), estimées entre 10 et 20 selon les auteurs, elles suscitent la controverse. Le linguiste américain Merritt Ruhlen avait déjà proposé de réunir toutes les langues en seulement trois familles: l'**eskimo-aléoute**, le **na-déné** et l'**amérinde**, cette dernière regroupant en une superfamille les langues d'Amérique, l'indo-européen et les langues d'Asie sous le nom de *eurasiatique*.

En somme, lorsque les linguistes se penchent sur l'origine des langues et des mécanismes du langage, ils se heurtent d'emblée au mystère même de l'histoire de l'humanité, ainsi que de son évolution à travers les millénaires.

Dernière mise à jour: 12 janvier, 2024

