

QUAND LES MOTS DISENT LES CHOSES

Une archéologie linguistique de la dyade mère-enfant

Julien d'Huy
Chercheur associé au LAS

INTRODUCTION

Dans *Les structures fondamentales des sociétés humaines*, Bernard Lahire propose une relecture ambitieuse des inégalités entre les sexes à partir d'une donnée anthropologique structurante, mais souvent négligée : la dyade femme-enfant. Selon lui, l'ancrage biologique de cette relation –, grossesse, accouchement, allaitement, soin rapproché d'une progéniture altricielle – aurait davantage contribué à façonner la répartition sexuée des tâches, des espaces et des pouvoirs dans les sociétés humaines que la seule domination masculine ou les rapports de force sociaux. Une grande part des inégalités durables entre les sexes, ultérieurement codifiées par des systèmes symboliques, religieux ou politiques, trouverait son origine dans cette configuration relationnelle originelle.

Le scénario ainsi proposé, nourri par les données de la biologie évolutive, de l'éthologie et de la sociologie, reste cependant partiel tant qu'on ne peut pas reconstruire la façon dont cette dyade mère-enfant était perçue, nommée et structurée par les premiers *Homo sapiens*. En effet, entre les contraintes biologiques et leur inscription dans des systèmes sociaux, se déploie un espace culturel où se formulent les représentations, se construisent les rôles, se fixent les normes. Si Lahire reconnaît ce maillon entre nature et culture, il le laisse en grande partie en suspens et n'en approfondit pas l'étude.

Le présent article vise à explorer cette zone intermédiaire à travers une méthode aréologique appliquée à la linguistique. L'objectif est de tester l'existence d'une structuration ancienne et géographiquement marquée de certaines formes

phonotactiques associées aux concepts de « mère » et de « sein ». Je soutiendrai l'hypothèse que certains lexèmes fondamentaux, tels que [na]/« mère », [mu] ou [amu]/« sein », ont pu se diffuser en même temps que les premières migrations d'*Homo sapiens*, et sont ainsi devenus des marqueurs résiduels de cette diffusion.

Dans une perspective plus large, des auteurs tels que Johanna Nichols (1992) ou moi-même (D'HUY, 2024 ; 2025a et b) avons pu souligner la capacité des distributions linguistiques à révéler des dynamiques anciennes de peuplement et de contact, tandis que Lyle Campbell (2004) rappelle les limites de l'inférence culturelle fondée uniquement sur la reconstruction lexicale. Les approches contemporaines insistent sur la nécessité de croiser les données linguistiques avec celles de l'archéologie et de l'anthropologie — une prudence déjà mise en avant par Leonard Bloomfield (1933) dans son exigence de démontrer strictement les relations génétiques. Ainsi, l'archéologie linguistique vise moins à reconstruire des « civilisations » complètes qu'à dégager, avec précaution, ce que les langues peuvent réellement révéler des structures et pratiques des sociétés passées.

Cette approche s'inscrit dans la tradition d'une « archéologie linguistique », selon les termes d'Edward Sapir¹, pour qui, « le mot suppose la chose connue ». La capacité à reconstruire un mot permettrait ainsi d'inférer l'existence passée de l'objet ou du concept qu'il désigne. La méthode prend racine dans la linguistique comparée du XIX^e siècle, lorsque des auteurs comme Franz Bopp² et August Schleicher³ ont montré que les correspondances systématiques entre langues permettaient de reconstruire des formes plus anciennes ; cette base méthodologique a ensuite consolidée par la découverte des lois phonétiques. Christopher Ehret⁴ considère que la reconstruction systématique d'un mot à l'aide de correspondances phonétiques régulières peut constituer une « preuve dure » en sciences historiques. Les mots peuvent alors être traités comme des artefacts culturels à part entière, au même titre que les outils ou les sépultures. Ainsi, la reconstruction du lexique proto-indo-européen permet d'esquisser un monde pastoraliste doté d'une culture de la roue et du cheval⁵. Toutefois, comme le rappelle Paul Heggarty⁶, cette démarche reste fragile : un mot peut désigner une

1. SAPIR, 1916.

2. BOPP, 1816.

3. SCHLEICHER, 1861.

4. EHRET, 2011.

5. MALLORY & ADAMS, 1997 et 2006 ; ANTHONY, 2007.

6. HEGGARTY, 2014.

notion plus ancienne, dérivée, ou symbolique. Par exemple, le terme « roue » pourrait dériver de notions telles que le cycle ou la circularité, sans faire référence à l'objet technique lui-même. Ce constat invite à considérer avec circonspection toute inférence linéaire entre lexique et culture matérielle.

C'est donc en s'armant de prudence que la présente étude se propose d'examiner la possibilité que les premières formes sonores associées aux concepts de maternité et d'allaitement aient laissé des traces dans les langues du monde. Cette hypothèse sera testée selon une perspective aréologique, notamment développée en linguistique par Johanna Nichols⁷ ou par l'auteur de cet article⁸, en analysant la répartition géographique des traits culturels et linguistiques pour en inférer l'origine, la diffusion et les transformations.

MATÉRIEL

J'ai utilisé les données de la base Lexibank⁹ auxquelles j'ai ajouté, pour les mots référant au « sein », les données de l'Automated Similarity Judgment Program¹⁰ – les deux bases de données fournissent des formes phonétiques alignées pour des milliers de langues à travers le monde. Quatre traits phonotactiques ont été sélectionnés :

- Pour « mère » (« *mother* ») : [n]/[ŋ] et [na]/[ɲa] à l'initial (ainsi que les variantes tonales et diacritées : [n.a], [ná], [nà], [nä], [ña], [ná], [nǎ], [nā], [nǎ́], [ñá], [næ], [nɑ], [ɲä]).
- Pour « sein » (« *breast* ») : [mu] et [amu].

Le choix des traits phonotactiques étudiés ([n]/[na] associés à « mère », [mu]/[amu] associés à « sein ») repose sur leur récurrence élevée à travers les langues du monde. Plusieurs études récentes ont ainsi montré que les sons nasaux, comme [ɲa] et [na] sont des marqueurs fiables des termes servant à désigner la mère¹¹, à la différence de la consonne [m] placée à l'initiale. Une telle association son/sens ne se retrouve pas pour d'autres termes semblables, comme ceux servant à désigner le père. De plus, cette corrélation va à l'encontre de la tendance enfantine à commencer les mots par une séquence consonne labiale/voyelle/consonne coronale,

7. NICHOLS, 1992 et 1994.

8. D'HUY, 2024, 2025a et 2025b.

9. LIST *et al.*, 2022, 2023 : <https://lexibank.clld.org/>.

10. WICHMANN *et al.*, 2022 : <https://asjp.clld.org/>.

11. JOHANSSON *et al.* 2020 ; IIDA & AKITA, 2024.

la production des sons labiaux comme [b], [p], [m], [f] et [v] étant plus facile que celle des sons coronaux pour les nourrissons et l'initiation des mouvements plus facile que leur continuation¹². Enfin, très peu de langues utilisent des sons nasaux vélares dans les termes parentaux, ce qui contraste fortement avec leur fréquence relativement élevée dans l'ensemble des termes pour « mère¹³ ». L'association privilégiée du mot « mère » et d'une nasale d'ouverture est donc suffisamment intrigante pour être explorée. De même, à l'échelle mondiale, les termes désignant le mot « sein » paraissent liés préférentiellement aux sons [muma]¹⁴, [m] et [u]¹⁵, ou encore à la combinaison [nasale +voyelle] (m)¹⁶.

Le corpus étudié se compose de 2959 langues pour « mère » et 7322 langues pour « sein », couvrant l'ensemble des continents habités. En complément, j'ai consulté les occurrences colexifiées « sein »/« lait » dans la base CLICS3¹⁷.

Les données ont été extraites sous forme de fichiers CSV, puis normalisées selon l'alphabet phonétique international (API). L'occurrence d'un trait a été codée « 1 » si la forme phonétique du *gloss*¹⁸ contenait la séquence ciblée (en position initiale pour « mère », en position initiale, médiane et/ou finale pour « sein ») et 0 dans le cas contraire. Ce codage binaire, bien qu'utile pour l'analyse, ne reflète pas la complexité réelle des usages lexicaux ou des contextes sémantiques. Il ne permet pas de pondérer l'importance lexicale ou phonologique réelle des dyades dans chaque langue – ce qui ne relève pas, en l'occurrence, des objectifs de cette étude. Les coordonnées géographiques de chaque langue ont été associées via les métadonnées des bases.

12. MACNEILAGE & DAVIS, 2000.

13. IIDA & AKITA, 2024.

14. WICHMANN *et al.*, 2010.

15. BLASI *et al.*, 2016.

16. JOHANSSON *et al.*, 2020.

17. RZYMSKI *et al.*, 2020.

18. Dans le présent article, le terme *gloss* est conservé dans son acception anglophone courante dans les bases de données lexicales comparatives (telles que Lexibank ou ASJP). Il désigne une étiquette sémantique standardisée permettant d'aligner et de comparer les formes lexicales issues de langues très diverses autour d'un même concept de référence (par exemple, *mother*, *breast*, etc.). Faute d'un équivalent strictement établi et stabilisé en français, et afin de rester cohérent avec la terminologie utilisée dans les corpus et les outils mobilisés, l'usage du terme *gloss* permet ici de maintenir la précision méthodologique et l'interopérabilité avec les ressources linguistiques internationales.

MÉTHODE

L'indice de Moran I a été calculé à partir des matrices de distance entre langues, en considérant la similarité des valeurs du trait dans les classes de distance. Une valeur significative de $I > 0$ indique un regroupement spatial du trait, une valeur < 0 indique une dispersion. Le logiciel utilisé pour ces analyses spatiales est PAST version 4.11¹⁹, dont l'algorithme s'appuie sur les travaux de Louis et Pierre Legendre²⁰.

J'ai également calculé la structure spatiale de chaque dyade à l'aide de l'indice de Moran I, mais, pour éviter les biais dus à une surreprésentation généalogique ou régionale, j'ai utilisé une matrice de poids fondée sur les 20 langues géographiquement les plus proches de chaque langue étudiée. Chaque test a été répété avec 1000 permutations aléatoires afin d'estimer la significativité du signal spatial indépendamment des regroupements familiaux ou historiques.

Dans un second temps, l'aire de diffusion de toutes les langues étudiées a été représentée sous forme de grilles régulières ($2^\circ \times 2^\circ$) afin de procéder à des analyses aréologiques de type spatial. Ce pas de grille correspond à une résolution d'environ 50 000 km² à l'équateur, ce qui garantit une couverture homogène tout en conservant un nombre suffisant d'observations par cellule dans la majorité des régions. La fréquence observée des traits a été comparée à leur fréquence attendue à l'aide de z-scores binomiaux. Ceux-ci ont été calculés pour chaque cellule selon la formule $z = (k - np) / \sqrt{np(1-p)}$, où k est le nombre de langues porteuses du trait dans la cellule, n le nombre total de langues dans la cellule, et p la fréquence globale de la dyade son/sens étudiée. Cela permet d'identifier les foyers de sur- ou sous-représentation en prenant en compte la densité géographique des langues, ce qui est fondamental dans la détection de structures aréales non aléatoires. Des tests complémentaires ont été menés avec des maillages plus fins ($1^\circ \times 1^\circ$) ou plus larges ($4^\circ \times 4^\circ$), sans modification significative des foyers observés. Cette stabilité interéchelle renforce la fiabilité des motifs aréologiques identifiés. Toutefois, dans les zones à très faible densité de langues, les z-scores doivent être interprétés avec prudence : une cellule isolée peut engendrer des valeurs extrêmes non représentatives.

Des permutations aléatoires ont été effectuées par tirage sans remise à l'aide de scripts Python, en maintenant constante la fréquence globale des traits. À chaque itération (n=1000), les langues codées « 1 » ont été redistribuées aléatoirement

19. HAMMER *et al.*, 2001.

20. LEGENDRE & LEGENDRE, 1998. Paramètres standards.

à travers l'ensemble du corpus, tout en conservant leur position géographique. Les z-scores ont été recalculés pour chaque simulation afin de générer une distribution de référence par cellule. Cette distribution simulée offre un intervalle de confiance empirique et permet d'évaluer la significativité des foyers de sur- ou sous-représentation identifiés dans les données réelles.

Les distances orthodromiques, c'est-à-dire les plus courtes distances entre deux points à la surface du globe (en tenant compte de sa courbure), ont été calculées à l'aide de la formule de Haversine, couramment utilisée en géodésie. Pour estimer la proximité spatiale entre deux traits linguistiques (par exemple, une forme phonotactique associée au mot « mère » et une autre au mot « sein »), la méthode a consisté à calculer, pour chaque langue présentant le trait A, la distance à la langue la plus proche présentant le trait B. La moyenne de ces distances minimales fournit un indicateur synthétique de la proximité géographique des distributions respectives des deux traits. En complément, une analyse de cooccurrence spatiale a été menée en superposant les indices géographiques des langues sur une grille régulière de $2^\circ \times 2^\circ$ de latitude et de longitude couvrant l'ensemble du globe. Le recoupement des cellules de cette grille, dans lesquelles au moins une langue présentait le trait A et au moins une autre le trait B, permet d'identifier les zones de convergence spatiale entre les deux distributions. Ces mesures croisées fournissent ainsi une double lecture – métrique et cartographique – des convergences spatiales observées. Celles-ci sont susceptibles de signaler une codiffusion ancienne, mais nécessitent des vérifications d'ordre généalogique et typologique.

RÉSULTATS

Les traits [n] et [ŋ] au début des mots signifiant « mère » sont associés à une autocorrélation spatiale significative (Indice de Moran, $p < 0.05$) (figure 1), avec des regroupements clairs dans l'hémisphère sud (figures 2 et 3). En prenant en compte les 20 langues les plus proches, l'indice de Moran I est d'environ 0,118, avec une p-value $< 0,001$, corroborant cette autocorrélation spatiale. La carte des z-scores donne à voir des foyers de surreprésentation en Afrique de l'Est, en Asie du Sud, dans l'Insulinde et la Mélanésie (figure 4). Ces résultats indiquent que ces traits ne sont pas répartis aléatoirement, mais suivent possiblement une logique diffusionniste.

La forme [na] – et ses variantes ([ná], [nä], etc.) – montre une structuration plus diffuse et globalement non significative (figures 5 et 6), mais avec des foyers localisés compatibles avec les routes de migration initiales d'*Homo sapiens* (figure 7). L'indice de Moran correspondant (figure 8) oscille autour de zéro, confirmant

l'absence d'autocorrélation spatiale significative à l'échelle globale. Cependant, en prenant en compte les 20 langues les plus proches, l'indice observé est d'environ 0,128, avec une p -value $< 0,001$, traduisant une structure régionale cohérente du lexique. La distribution des distances géographiques entre les langues partageant la forme [na] (figure 9) montre également une forte concentration de comparaisons à courte distance, sans correspondance linéaire avec la fréquence du trait, ce qui confirme un effet localisé, mais non systématique. Cette configuration suggère soit des développements indépendants, soit un phénomène d'archaïsme local conservé dans certaines régions, soit une diffusion initiale ancienne partiellement estompée par des processus de contact et de remplacement.

Pour les mots désignant le « sein », les formes [mu] et [amu] présentent également une structuration spatiale nette (figures 10 et 11).

L'analyse de l'indice de Moran pour [mu] (figure 12) révèle une autocorrélation spatiale significative à très courte portée, avec une décroissance rapide. Cela suggère une structuration locale robuste, mais peu étendue, compatible avec une transmission verticale dans des régions isolées, autrement dit, un effet d'isolation par la distance. Le calcul, en prenant en compte les 20 langues les plus proches spatialement, corrobore cette analyse : l'indice de Moran I observé est d'environ 0,154, avec une p -value $< 0,001$, ce qui indique que les langues géographiquement proches tendent à se ressembler davantage que ce que l'on observerait par hasard. La carte des z-scores binomiaux associés à [mu] (figure 13) confirme cette interprétation : on observe une surreprésentation marquée de cette forme dans plusieurs régions précises, notamment l'Asie du Sud-Est insulaire, la Papouasie-Nouvelle-Guinée et la Mélanésie. Ces zones correspondent aux corridors côtiers empruntés par les premières vagues de peuplement humain hors d'Afrique, ce qui pourrait appuyer – avec prudence – l'hypothèse d'un transport ancien de la forme [mu] dans le cadre de ces migrations fondatrices. À l'inverse, les régions d'Afrique occidentale, d'Eurasie septentrionale et des Amériques présentent une sous-représentation statistique significative, qui pourrait résulter de phénomènes de remplacement ou d'érosion lexicale. Enfin, une concentration modérée de scores positifs à proximité de la Corne de l'Afrique, bien que plus diffuse, pourrait évoquer un foyer initial de la dyade ou une conservation locale antérieure aux premières dispersions hors d'Afrique – une hypothèse à traiter avec précaution, en l'absence de données archéolinguistiques directes.

L'analyse de la forme [amu] pour « sein » (figure 14) montre également une autocorrélation positive à très courte portée, suivie d'un effondrement rapide, typique là encore d'un effet de proximité géographique. Corrigé en prenant en compte les 20 langues les plus proches, l'indice de Moran I atteint environ 0,168

($p < 0,001$), confirmant une structuration géographique marquée. Toutefois, la diffusion de [amu] apparaît plus restreinte que celle de [mu], ce que met en évidence la carte des z-scores correspondante (figure 15). Seules quelques zones bien délimitées – notamment en Mélanésie, dans les îles Salomon et en Papouasie-Nouvelle-Guinée – affichent une surreprésentation significative. Toutefois, une lecture attentive de la carte montre une surreprésentation modérée, mais significative en Afrique, notamment dans la bande sahéenne et certaines zones de l’Afrique de l’Est. Cette élévation relative, bien moins marquée que dans les foyers extrêmes, pourrait correspondre à un « signal fossile » d’un foyer plus ancien et diffus. Associée à la présence affirmée de l’association [amu]/« mère » en Afrique, cette configuration suggère que ce continent pourrait avoir été un centre initial d’émergence ou de structuration de la dyade, suivi d’une dispersion progressive vers d’autres régions. Par la suite, la dyade aurait été partiellement masquée par des recompositions ultérieures.

Afin de tester la robustesse statistique des structures spatiales mises en évidence pour les différentes dyades étudiées, une procédure de permutation aléatoire a été appliquée, dans la continuité des méthodes décrites plus haut. Pour chacune des 1 000 itérations, les langues portant un trait donné (codées « 1 ») ont été redistribuées aléatoirement à travers l’ensemble du corpus, sans modification de leur fréquence globale, ce qui permet de générer une situation de référence conforme à l’hypothèse nulle d’une répartition aléatoire. À chaque permutation, les z-scores binomiaux locaux – utilisés précédemment pour détecter les foyers de sur- ou sous-représentation géographique – ont été recalculés, produisant ainsi une distribution simulée des valeurs attendues en l’absence de structuration spatiale. La comparaison entre les z-scores issus des données observées et ceux obtenus dans ces distributions permutées permet d’évaluer la significativité des motifs géographiques repérés, en distinguant les structures réelles des effets susceptibles d’émerger par hasard. Les figures 16 à 19 présentent les z-scores moyens des permutations avec une visualisation croisée par cellule (grille de $2^\circ \times 2^\circ$). Les résultats confirment que la structuration géographique observée pour les traits [n], [na], [mu] et [amu] ne peut s’expliquer par une simple répartition aléatoire : des foyers de surreprésentation significatifs persistent dans des régions clés, en particulier en Afrique orientale, en Asie méridionale, en Insulinde et en Mélanésie. Ce test de permutation renforce l’hypothèse d’une codiffusion ancienne de certaines dyades signifiant/signifié en lien avec la mère et le sein le long des axes de dispersion d’*Homo sapiens*.

Afin de tester l’hypothèse de cette codiffusion lexicale ancienne entre les dyades [n] et [na]/« mère », [mu] et [amu]/« sein », j’ai analysé leur distribution

géographique à l'échelle mondiale. Deux indicateurs ont été mobilisés : la distance moyenne en kilomètres entre chaque point d'un corpus et son plus proche voisin dans l'autre corpus, et le nombre de cellules géographiques partagées (grilles de 2°× 2° de latitude/longitude). Les comparaisons entre [n] et [mu], ainsi qu'entre [na] et [mu] (figures 20 et 21), montrent une proximité géographique étroite, avec des distances moyennes faibles (<450 km) et un recouvrement cellulaire élevé (> 85 cellules partagées). À l'inverse, les associations avec [amu] révèlent une dissociation spatiale marquée (distances moyennes > 1400 km, moins de 40 cellules partagées). Ces résultats suggèrent que [mu] partage avec [n] et [na] une aire de diffusion conjointe bien définie, probablement héritée d'un noyau lexical ancien, tandis que [amu] aurait émergé, évolué ou se serait conservé dans des contextes plus isolés ou secondaires.

J'ai ensuite précisé les régions géographiques correspondant à ces superpositions, en attribuant chaque cellule partagée à une aire culturelle ou continentale. L'analyse repose sur une grille régulière couvrant l'ensemble du globe, dans laquelle chaque cellule représente une surface de 2 degrés de latitude par 2 degrés de longitude. À l'équateur, cela correspond approximativement à 222 km×222 km, soit environ 50 000 km² par cellule, avec une surface qui diminue progressivement vers les pôles en raison de la convergence des méridiens. Cette cartographie permet d'identifier trois foyers principaux de cooccurrence (figure 22) : la Mélanésie, l'Australie et le Pacifique central. Dans cette analyse, « Mélanésie » recouvre la Papouasie-Nouvelle-Guinée, les îles Salomon, le Vanuatu et les archipels avoisinants, « Australie » désigne la masse continentale formée par le pays éponyme, et « Pacifique central » inclut des îles plus orientales, comme les Fidji ou Tonga. Aucune cooccurrence n'a été relevée en Afrique, en Eurasie ou dans les Amériques. Ces localisations très méridionales pourraient appuyer l'hypothèse d'une codification lexicale depuis l'Afrique via la route australe, conservée dans des régions longtemps restées à l'écart des grandes reconfigurations linguistiques.

DISCUSSION

Certains chercheurs ont avancé que les sons comme [ma] sont souvent associés à la mère, car ils pourraient être liés au babillage des bébés. Ce dernier privilégierait des sons simples et contrastés, comme une consonne nasale ([m]) suivie d'une voyelle basse ([a])²¹. Selon cette théorie, le bébé, sans pensée particulière, babille-

21. MURDOCK, 1959 ; JAKOBSON, 1960.

rait ; la mère, accueillie par son enfant avec le son « ma », aurait considéré qu'il lui est adressé, et, comme elle viendrait fréquemment au berceau en entendant ce son, l'enfant lui-même apprendrait à utiliser ces syllabes pour l'appeler²². Une théorie similaire lie la connexion [ma]/« mère » à l'allaitement : le son [m] serait un murmure anticipatoire produit par le bébé, tandis que le son [a] résulterait de l'ouverture de la bouche en préparation pour téter.

Cependant, une telle approche souffre de biais importants. Contrairement aux idées reçues, [ma] n'est pas particulièrement lié à la mère : il est utilisé de manière équivalente pour désigner la mère ou le père selon les langues, ce qui remet en question l'universalité de l'hypothèse liant [ma] à la maternité²³ et corrobore le fait, mis en évidence par diverses études, que la ressemblance attribuée entre le sens et la forme d'un signe est culturellement acquise²⁴.

Est-il alors possible d'articuler l'hypothèse selon laquelle une dyade [n]/[na] aurait désigné la mère chez les premiers *Homo sapiens*, tout en expliquant la dominance actuelle de [m]/[ma] pour référer au même terme à travers les langues ? Pourrait-on imaginer que les formes en [na] dérivent historiquement de formes en [ma] ? En phonétique historique, le passage de [m] (nasale bilabiale) à [n] (nasale alvéolaire) est bien attesté, notamment dans des contextes d'assimilation progressive ou régressive à des segments voisins²⁵. Ce type de mutation correspond à un recul du point d'articulation – de l'avant vers l'arrière de la cavité buccale –, ce qui s'accorde avec les processus articulatoires naturels les plus courants. À l'inverse, le passage de [n] vers [m], qui requiert une avancée du point d'articulation, est plus rare, sauf dans certains contextes très spécifiques, comme l'assimilation explicite à une consonne bilabiale suivante. Les changements de lieu d'articulation sont rarement orientés vers l'avant (c'est-à-dire de coronale à bilabiale) sans pression morphophonologique ou environnementale spécifique. En ce sens, la trajectoire [ma]>[na] est phonétiquement plausible et cohérente avec les modèles classiques d'évolution phonétique.

Cependant, si l'on observe que les formes en [ma] sont omniprésentes dans les langues du monde pour désigner la mère, les formes en [na] sont limitées à certaines zones restreintes de l'Ancien Monde (Afrique, Asie du Sud-Est, Océanie). Une simple transformation phonétique du type [m]>[n] ne suffit pas à expliquer

22. JESPERSEN, 1922, p. 154.

23. PASSMORE *et al.*, 2023.

24. IIDA & AKITA, 2024.

25. LASS, 1997 ; TRASK, 1996.

cette distribution. On peut alors avancer que la dyade [na], bien qu'aujourd'hui minoritaire, ait constitué une forme originelle dans les premières langues humaines, avant d'être progressivement remplacée par des innovations indépendantes menant à [ma]. Ainsi, la profusion de [ma] dans les langues modernes ne refléterait pas nécessairement une origine commune, mais plutôt un biais articulatoire universel en faveur des nasales bilabiales, particulièrement accessibles dans le babillage. Ce biais aurait pu effacer ou noyer les traces d'une forme initiale en [na], plus ancienne, mais moins stable.

De la même façon, le fait que les mots renvoyant à « poitrine » et « lait » sont associés à des groupes de sons [nasale] et [nasale+voyelle] a été expliqué par le fait que les sons nasaux sont couramment produits par les nourrissons lorsqu'ils sont allaités, l'obstruction de leur bouche empêchant la respiration et la production de sons oraux²⁶. Lyle Campbell note ainsi que : « Les similitudes entre ces mots dans diverses langues sont généralement attribuées à l'imitation des sons produits par les enfants lorsqu'ils têtent ou sucent²⁷. » Or, comme l'a montré cet article, l'association des mots renvoyant à « poitrine » et « lait » à des groupes de sons [nasale] et [nasale+voyelle] n'est pas universelle.

Par ailleurs, l'hypothèse d'un symbolisme sonore associant « lait » et « sein » aurait dû conduire à l'apparition de multiples colexifications de ces deux termes, c'est-à-dire l'usage d'un mot unique pour signifier ces deux choses différentes²⁸. Les termes seraient alors rapprochés par le son autant que par une relation métonymique. Or cette colexification n'est pas universelle. Matthias Urban²⁹ a noté que si la colexification entre les mots « sein » et « lait (maternel) » est extrêmement courante, elle paraît virtuellement absente en Eurasie continentale, où elle ne se trouverait qu'en basque et en ket. Une consultation du site *Clics3*³⁰ appuie ce constat, tout en rajoutant quelques rares occurrences dans le nord de l'Eurasie (figure 23).

Les résultats soutiennent l'idée que certaines formes sonores liées à la maternité et à l'allaitement ne relèvent pas uniquement de contraintes articulatoires universelles, mais pourraient résulter d'une stabilisation culturelle ancienne, transmise dans certaines aires linguistiques. Leur répartition géographique, leur

26. JOHANSSON *et al.*, 2020.

27. CAMPBELL, 2008, p. 88.

28. FRANÇOIS, 2008.

29. URBAN, 2012, p. 624-659.

30. *Clics3*, URL : <https://clics.clld.org/> ; RZYMSKI *et al.*, 2020

stabilité et leur association à des *gloss* comme « mère » et « sein » suggèrent une transmission culturelle ancienne. Par ailleurs, les deux termes sont liés, géographiquement et symboliquement, à travers la colexification « sein »/« lait ».

Plusieurs arguments appuient l'hypothèse d'une telle conservation. En se basant sur un ensemble de données de 107 langues eurasiennes, Johannes Dellert et ses collègues³¹ ont montré que si les voyelles sonores sont les sons les plus instables dans le temps, les labiales, les nasales et les latérales, présentes dans nos reconstructions, possèdent une stabilité presque complète. Par ailleurs, l'emploi de consonnes nasales vélares voisées paraît évoluer lentement et demeurer stable dans une même langue³². Enfin, concernant le mot « mère », la consonne en position initiale tend à demeurer particulièrement stable³³. De façon plus générale, plusieurs études ont mis en évidence l'importante stabilité des mots désignant le « sein³⁴ ».

L'alignement partiel entre ces foyers lexicaux et les hypothèses paléoanthropologiques sur les routes de migration humaines, se diffusant d'abord via le sud de l'Eurasie jusqu'au Sahul³⁵ renforce l'idée d'une codiffusion lexicale, tout en permettant de la dater. Le cas de [mu] est particulièrement révélateur : il se situe dans des régions culturellement isolées, loin des grands foyers de standardisation linguistique comme l'Eurasie continentale. Cela pourrait correspondre à une conservation de traits hérités, échappant aux grands mouvements d'unification ultérieurs. Il est à noter que contrairement aux travaux de John Bengston et Merritt Ruhlen³⁶, ce n'est pas ici l'universalité, mais la non-universalité des dyades, qui permet de leur supposer un âge paléolithique.

Le complexe de sens « mère »/« sein » aurait pu s'opposer à un autre, associé au père. S'il existe une relation entre [ta] et [pa] et les termes désignant le père³⁷, cette relation présente une diffusion moins australe que celle existant entre [ŋa] et [na] et les termes désignant la mère³⁸. On peut alors supposer l'existence d'une différenciation maximale entre les sons, qui aurait agi comme un « moteur »

31. DELLERT *et al.*, 2021.

32. DEDIU & CYSOUW, 2013 ; KAUFMAN *et al.*, 2021.

33. SCHEER & SÉGÉRAL, 2001 ; BENDJABALLAH & GAC, 2023.

34. (HOLMAN *et al.*, 2008 ; TADMOR *et al.*, 2010.)

35. BAE *et al.*, 2017.

36. BENGSTON, 1994 ; RUHLEN, 2007.

37. PASSMORE *et al.*, 2023.

38. LIST *et al.*, 2022.

phonologique dans la construction des termes parentaux et, de façon plus générale, pour l'établissement des genres. La culture fonctionnerait ainsi comme un maximisateur de différences, à partir d'un noyau primaire mère-enfant (renvoyant au « lien particulièrement fort noué entre la mère et l'enfant »), et l'adjonction périphérique du père (dont le lien avec l'enfant est « incroyablement plus lâche³⁹ »).

Cependant, des études complémentaires seront nécessaires pour écarter définitivement l'hypothèse d'une convergence articulatoire partiellement indépendante du sens⁴⁰ à l'intérieur de nos corpus. Par ailleurs, la stabilité phonotactique ne garantit pas une continuité sémantique stricte : les mots peuvent changer de référents tout en conservant leur forme. Enfin, les bases de données linguistiques utilisées sont par nature incomplètes.

Malgré ces limites, la méthode aréologique permet de dépasser les simples comparaisons lexicostatistiques en introduisant une dimension spatiale et historique. Elle suggère que l'étude du langage peut, lorsqu'elle est soigneusement encadrée, offrir un accès indirect aux représentations les plus anciennes de l'humanité.

CONCLUSION

Cette étude a mis en évidence une structuration géographique ancienne et non aléatoire de certains traits phonotactiques associés aux concepts de « mère » et de « sein ». Les résultats suggèrent une codiffusion lexicale pour certaines de ces formes (notamment [n], [na], [mu]) le long des premières routes de migration humaine hors d'Afrique.

Cette coïncidence entre phonotactique, *gloss*, et diffusion géographique suggère l'existence possible d'un noyau lexical partagé, compatible avec l'hypothèse d'une unité culturelle ancienne autour de la dyade mère-enfant. En cela, elle offre un soutien partiel à la thèse de Bernard Lahire, tout en appelant à une exploration plus approfondie des substrats linguistiques du social.

Loin de prétendre reconstituer un protolangage unique, cette étude met plutôt en lumière les traces résiduelles de traits culturels anciens à travers leur expression linguistique, invitant à considérer les mots comme des artefacts au même titre que les outils, les sépultures ou les mythes.

39. LAHIRE, 2023, p. 769.

40. JAKOBSON, 1960.

ANNEXES : FIGURES

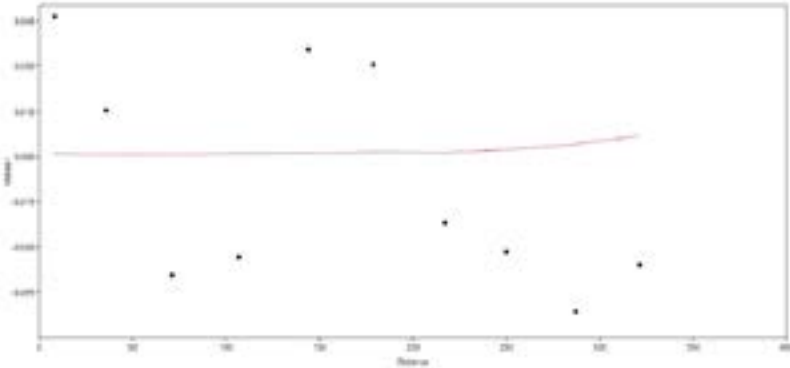


Figure 1. Analyse de l'autocorrélation spatiale pour les sons [n] et [ŋ] en position initiale des mots signifiant « mère », en fonction des bandes de distance.

>0=clustering ; <0=dispersion. La ligne rouge indique le seuil de significativité ($p < 0.05$).
 Une valeur significative à courte distance suggère que ces traits sont regroupés géographiquement, indiquant une diffusion non aléatoire.



Figure 2. Carte mondiale des langues utilisant le son [n] au début du mot « mère ».

La forte densité dans l'hémisphère sud (Afrique, Asie du Sud, Océanie) soutient l'idée d'une transmission ancienne liée aux premières migrations humaines.



Figure 3. Répartition des mots pour « mère » commençant par [ɲ].

Ce son plus rare montre une concentration régionale en Papouasie et dans l'Insulinde, possiblement due à une préservation locale d'un archaïsme lexical ou un développement indépendant.

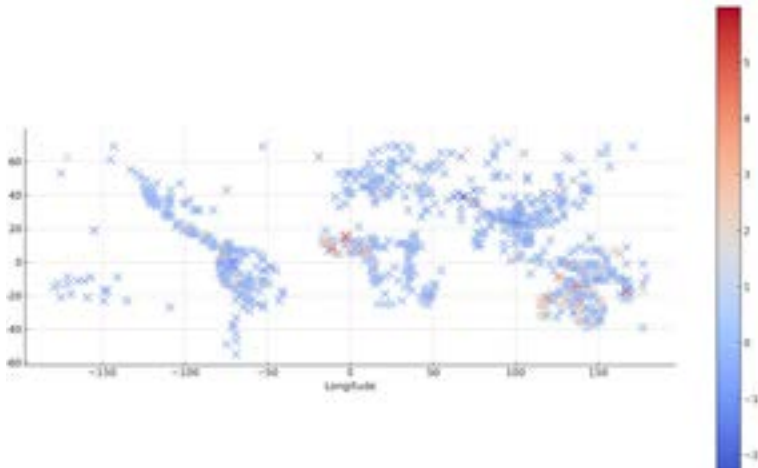


Figure 4. Z-scores pour la dyade [n]/[ɲ] à l'initial des mots « mère » : cette carte mesure la sur- ou sous-représentation par rapport à une attente statistique.

Chaque cellule représente une grille de 2° x 2° de latitude et de longitude. Les zones rouges correspondent à une surreprésentation significative du trait, tandis que les zones bleues indiquent une sous-représentation par rapport à une répartition aléatoire. Les foyers les plus significatifs apparaissent en Afrique de l'Est et en Asie du Sud-Est, zones associées à la diffusion initiale d'Homo sapiens. Cellules rouges foncées ($z > 2$) : Surreprésentation significative du trait par rapport à l'attendu ($p < 0,05$) ; Cellules oranges/jaunes (z entre 0 et 2) : Léger excès du trait, non significatif ; Cellules claires/blanches ($z \approx 0$) : Concordance avec la fréquence attendue sous hypothèse de distribution aléatoire ; Cellules bleues pâles à foncées ($z < 0$) : Sous-représentation du trait, potentiellement significative en bleu foncé ($z < -2$) ; Cellules grises/vides : Absence de données linguistiques dans cette zone.



Figure 5. Répartition mondiale des mots « mère » commençant par [na].

Cette forme présente une structure plus diffuse, mais partiellement superposable à [n], ce qui pourrait indiquer un développement commun ou une diversification secondaire.



Figure 6. Carte des occurrences du son [ɲa] dans les mots pour « mère ».

Très rare, cette forme se concentre en Mélanésie, ce qui peut signaler un vestige phonotactique ancien préservé ou un développement indépendant à partir des mots « mères » possédant [n] à l'initial.

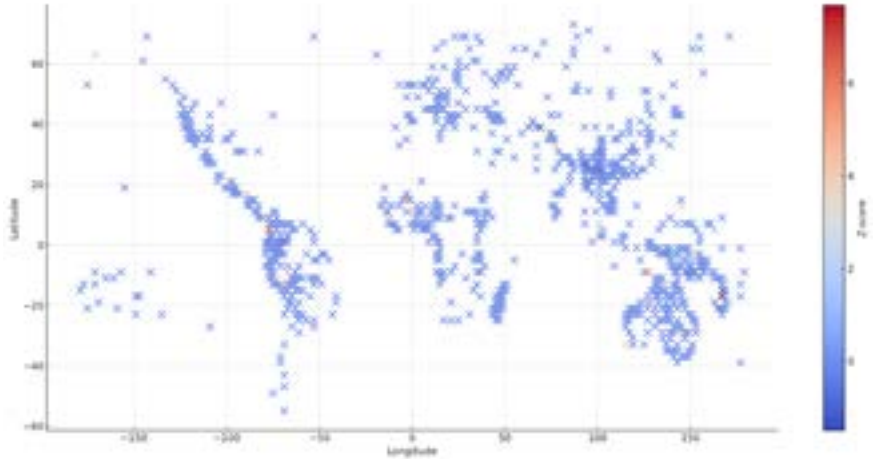


Figure 7. Z-scores pour les formes [na]/[ɲa].

Des foyers de surreprésentation sont visibles en Afrique de l'Est, en Asie du Sud, en Mélanésie et en Amazonie. Ces zones correspondent à des régions de continuité linguistique ancienne ou de faible contact.

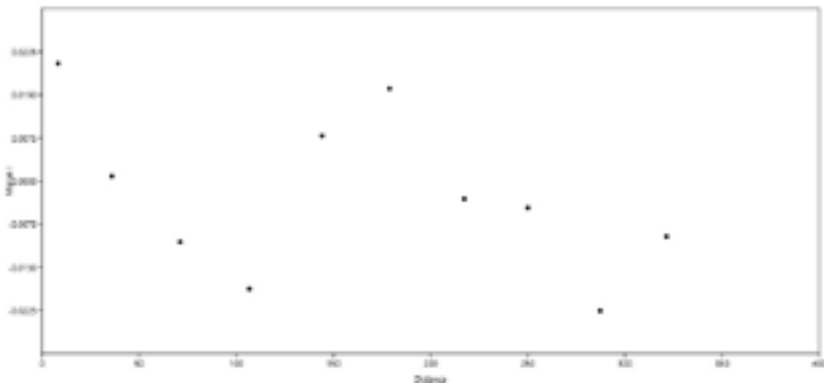


Figure 8. Autocorrélation spatiale pour [na]/[ɲa].

L'indice de Moran proche de zéro indique que, globalement, ces formes ne présentent pas de structuration spatiale forte, suggérant une diffusion moins systématique.

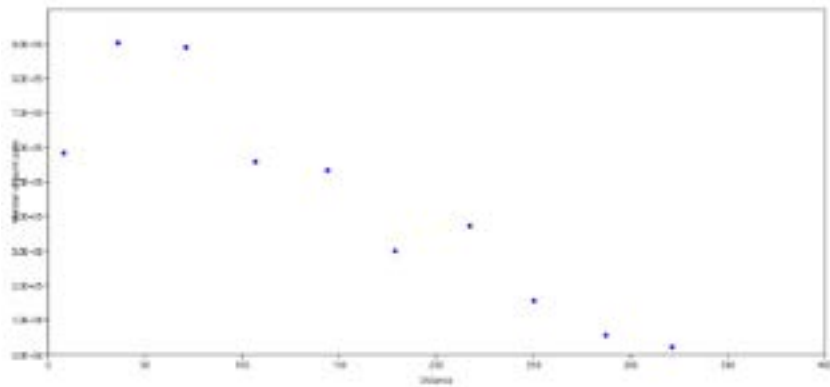


Figure 9. Distances géographiques entre langues partageant la forme [na]/[ɲa].

Une forte densité à courte distance, sans schéma cohérent à grande échelle, évoque des foyers conservés localement plutôt qu'un processus global.



Figure 10. Répartition des langues contenant [mu] dans les mots pour « sein ».

Les foyers sont concentrés en Asie du Sud-Est et en Océanie, zones empruntées par les premières vagues de peuplement humain.



Figure 11. Carte des langues utilisant [amu] pour désigner le « sein ».

Sa diffusion est plus restreinte, concentrée en Mélanésie, Australie et quelques points en Afrique de l'Est, ce qui suggère un marqueur résiduel plus ancien encore.

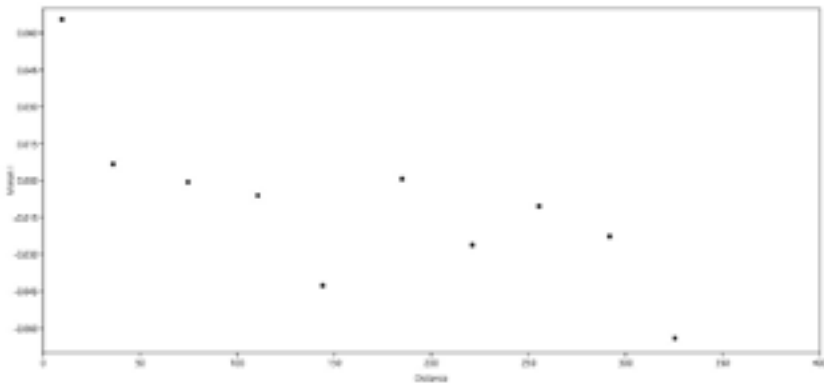


Figure 12. Autocorrélation spatiale pour [mu] dans les mots pour « sein ».

La structuration est forte à très courte distance, ce qui témoigne d'une transmission verticale dans des zones isolées.

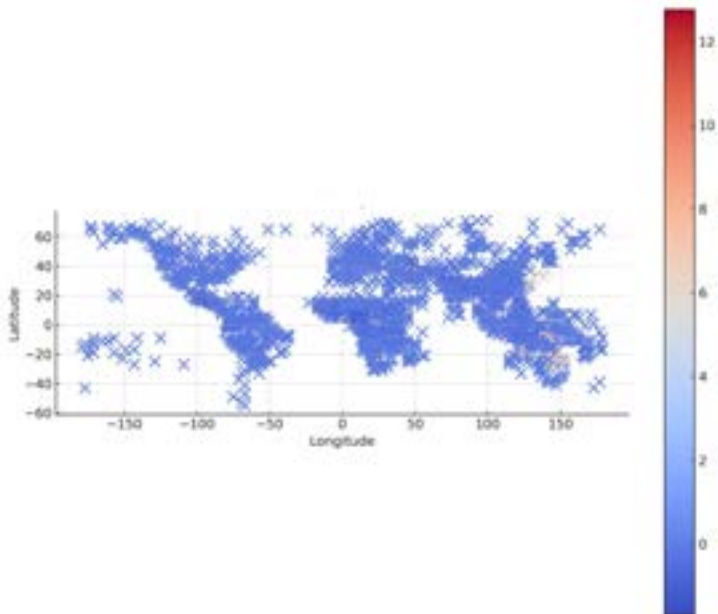


Figure 13. Z-scores pour [mu].

La surreprésentation statistique se manifeste nettement dans les zones australes, alors que l’Eurasie et les Amériques sont sous-représentées, ce qui indique une ancienne diffusion conservée dans des marges culturelles.

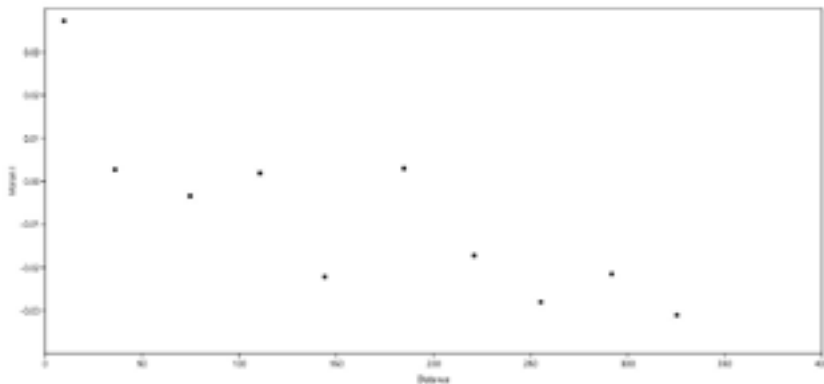


Figure 14. Indice de Moran pour [amu].

Une structuration positive à courte distance suivie d’un effondrement rapide suggère que ce trait est maintenu dans des microrégions isolées, ce qui est typique des archaïsmes linguistiques.

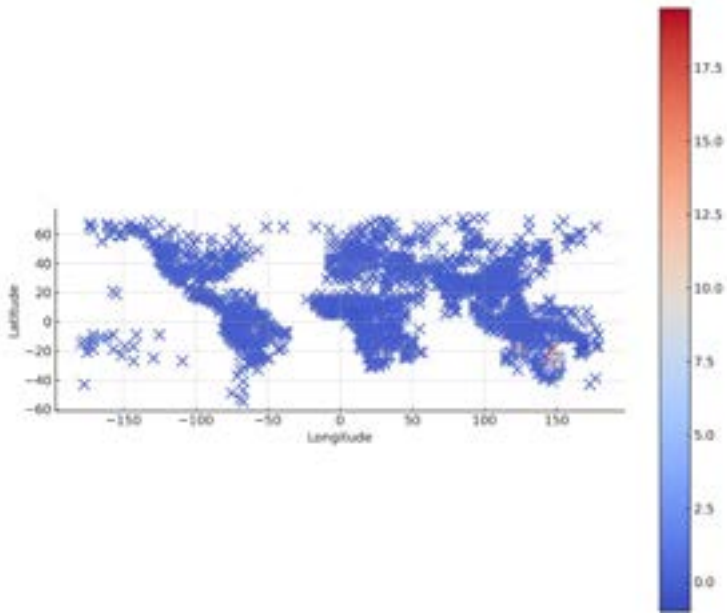


Figure 15. Z-scores pour [amu].

Les foyers de surreprésentation se concentrent en Mélanésie et dans l'Insulinde. L'absence sur d'autres continents suggère un trait résiduel, très ancien et peu sujet à expansion récente.

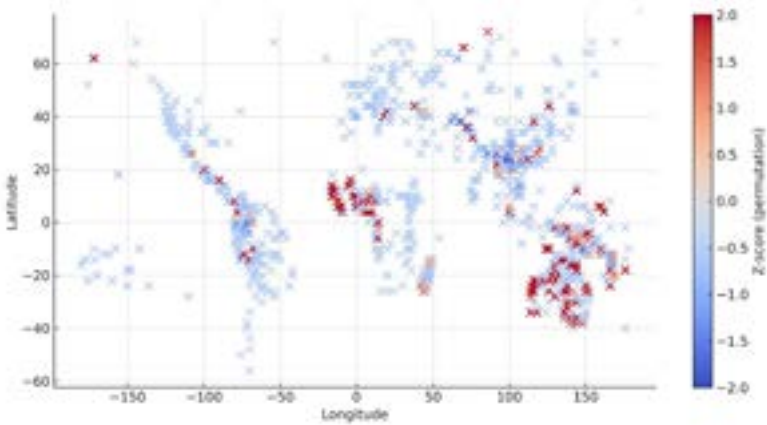


Figure 16. Résultats des permutations aléatoires pour [n]/[ŋ].

Même après 1 000 redistributions aléatoires, les foyers observés persistent, ce qui confirme que leur structuration spatiale est significative et non due au hasard.

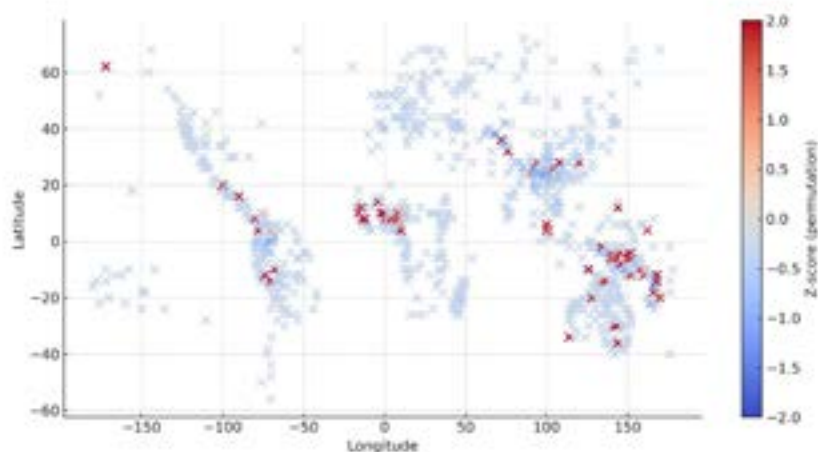


Figure 17. Résultats des permutations aléatoires pour [na]/[ɲa].

Bien que les foyers soient visibles, ils sont moins nets et plus diffus que pour [n], ce qui suggère une évolution lexicale plus récente ou une stabilité moindre.

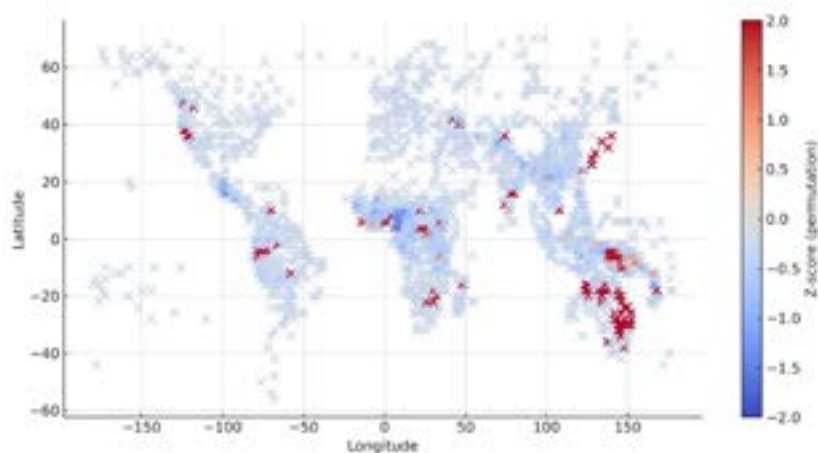


Figure 18. Résultats des permutations aléatoires pour [mu].

Les foyers observés en Asie du Sud-Est et Mélanésie persistent au-delà des attentes aléatoires, renforçant l'idée d'un archaïsme lexical diffusé avec les migrations humaines.

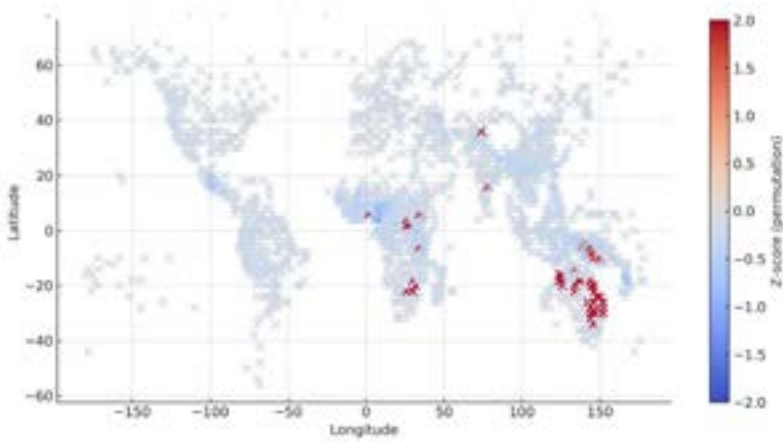


Figure 19. Résultats des permutations aléatoires pour [amu].

Malgré sa rareté, la structuration reste stable dans son aire de diffusion, renforçant l'hypothèse d'un trait phonotactique ancien conservé localement.

Comparaison	Distance moyenne (km)
[n] → [mu]	427
[n] → [amu]	1 447
[na] → [mu]	420
[na] → [amu]	1 779

Figure 20. Distance moyenne entre langues partageant les dyades étudiées.

Les combinaisons [n]/[na] avec [mu] affichent les distances les plus courtes (<450 km), indiquant une codiffusion potentielle. Les distances avec [amu] sont bien plus grandes.

Comparaison	Cellules partagées
[n] ∩ [mu]	92
[n] ∩ [amu]	36
[na] ∩ [mu]	87
[na] ∩ [amu]	28

Figure 21. Nombre de cellules géographiques partagées entre traits dans une grille 2°× 2°.

Plus de 85 cellules communes entre [n]/[na] et [mu], contre moins de 40 pour [amu], indique une aire de recouvrement significative uniquement pour [mu].

Région	[n] ∩ [mu]	[n] ∩ [amu]	[na] ∩ [mu]	[na] ∩ [amu]
Mélanésie	12	7	5	1
Australie	4	3	2	2
Pacifique central	1	0	0	0

Figure 22. Répartition régionale des cellules partagées pour la Mélanésie, l’Australie et le Pacifique central.

Les cooccurrences entre traits « mère » et « sein » se concentrent dans trois zones (Mélanésie, Australie, Pacifique central), toutes situées sur les routes méridionales de dispersion humaine.



Figure 23. Carte mondiale des colexifications entre « sein » et « lait ».

Ce lien est fréquent partout dans le monde sauf en Eurasie continentale, ce qui indique une séparation sémantique propre à cette région culturelle.

BIBLIOGRAPHIE

BAE Christopher J., DOUKA Katerina & PETRAGLIA Michael D., 2017, “On the origin of modern humans: Asian perspectives”, *Science* vol. 358, n° 6368, DOI : 10.1126/science.aai9067.

DE BEAUNE Sophie A., 2022, *Préhistoire intime*, Gallimard, Paris.

BEEN Ella & BARZILAI Omry, 2024, “Neandertal burial practices in Western Asia: How different are they from those of the early *Homo sapiens*?”, *L’Anthropologie, Sépultures paléolithiques et mésolithiques*, vol. 128, n° 3, DOI : 10.1016/j.anthro.2024.103281.

- BENDJABALLAH Sabrina & LE GAC David, 2023, “The Acoustics of Word-Initial and Word-Internal Voiced Stops in Somali” in *Journal of the International Phonetic Association*, vol. 53, n° 3, p. 644-81, DOI : 10.1017/S0025100321000281.
- BENGSTON John D. & RUHLEN Merritt, 1994, “Global etymologies” in RUHLEN Merritt, *The origin of language: tracing the evolution of the mother tongue*, p. 277-366, New York J. Wiley.
- BLASI Damián E., WICHMANN Søren, HAMMARSTRÖM Harald, STADLER Peter F. & CHRISTIANSEN Morten H., 2016, “Sound–meaning association biases evidenced across thousands of languages” *Proceedings of the National Academy of Sciences*, vol. 113, n° 39, DOI : 10.1073/pnas.1605782113.
- BOPP Franz, 1816, *Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*, Andreätschen, Francfort-sur-le-Main, 312 p.
- CAMPBELL Lyle, 2008, “What Can We Learn about the Earliest Human Language by Comparing Languages Known Today?” in CAMPBELL Lyle & POSER William J. (dir.), *Language Classification: History and Method*, p. 79-111, Cambridge University Press, Cambridge, DOI : 10.1017/CBO9780511486906.
- DEDIU Dan et CYSOUW Michael, 2013, “Some structural aspects of language are more stable than others: a comparison of seven methods”, *Plos One*, vol. 8, n° 1, DOI : 10.1371/journal.pone.0055009.
- DELLERT Johannes, JOHANSSON Niklas Erben, FRID Johan et CARLING Gerd, 2021, “Preferred sound groups of vocal iconicity reflect evolutionary mechanisms of sound stability and first language acquisition: evidence from Eurasia”, *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 376, n° 1824, DOI : <https://doi.org/10.1098/rstb.2020.0190>.
- EHRET Christopher, 2012, “Linguistic Archaeology”, *African Archaeological Review*, vol. 29, n° 2-3, p. 109-30, DOI 10.1007/s10437-012-9116-x.
- FRANÇOIS Alexandre, 2008, “Semantic Maps and the Typology of Colexification: Intertwining Polysemous Networks across Languages” in VANHOVE Martine (dir.), *From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations*, p. 163-215, Amsterdam, John Benjamins Publishing Company, DOI : 10.1075/slcs.106.09fra.
- GALOR Oded, ÖZAK Ömer & SARID Assaf, 2020, “Linguistic traits and human capital formation”, *AEA Papers and Proceedings*, 110, p. 309-13,

Nashville, American Economic Association, <https://www.aeaweb.org/articles?id=10.1257/pandp.20201069>.

HAMMER Øyvind & HARPER David AT, 2001, “Past: paleontological statistics software package for education and data analysis”, *Palaeontologia electronica*, vol. 4, n° 1, p. 1-9.

HEGGARTY Paul, 2015, “Prehistory through language and archaeology” in BOWERN Claire & BETHWYN Evans (dir.), *The Routledge Handbook of Historical Linguistics*, p. 598-628, Routledge, Londres & New York.

HENRY-GAMBIER Dominique, 1986, « Étude des os d’enfants du gisement aurignacien de Cro-Magnon, Les Eyzies (Dordogne) », *Bulletins et mémoires de la société d’anthropologie de Paris* 3, n° 1, p. 13-25, DOI : <https://doi.org/10.3406/bmsap.1986.1590>.

JAKOBSON Roman, “Why ‘Mama’ and ‘Papa’?” in WERNER Heinz, KAPLAN Bernard & WAPNER Seymour (dir.), *Perspectives in psychological theory. Essays in honor of Heinz Werner. Bernard Kaplan and Seymour Wapner*, International U.P., New York, p. 124-34.

JESPERSEN Otto, 1922, *Language, Its Nature, Development and Origin*, G. Allen & Unwin, Londres.

JOHANSSON Niklas Erben, ANIKIN Andrey, CARLING Gerd & HOLMER Arthur, 2020, “The Typology of Sound Symbolism: Defining Macro-Concepts via Their Semantic and Phonetic Features” in *Linguistic Typology* vol. 24, n° 2, p. 253-310, DOI : [10.1515/lingty-2020-2034](https://doi.org/10.1515/lingty-2020-2034).

KAUHANEN Henri, GOPAL Deepthi, GALLA Tobias & BERMÚDEZ-OTERO Ricardo, 2021, “Geospatial distributions reflect temperatures of linguistic features”, *Science Advances*, vol. 7, n° 1, DOI : [10.1126/sciadv.abe6540](https://doi.org/10.1126/sciadv.abe6540).

LAHIRE Bernard, 2023, *Les structures fondamentales des sociétés humaines*, La Découverte, Paris.

LASS Roger, 1997, *Historical linguistics and language change*, Cambridge University Press, Cambridge.

LEGENDRE Pierre & LEGENDRE Louis, 1998, *Numerical ecology*, Elsevier, Amsterdam, 1006 p.

LIST Johann-Mattis, FORKEL Robert, GREENHILL Simon J., RZYMSKI Christoph, ENGLISH Johannes & GRAY Russell D., 2022, “Lexibank, a Public Repository of Standardized Wordlists with Computed Phonological and Lexical Features”, *Scientific Data*, vol. 9, n° 1, DOI : [10.1038/s41597-022-01432-0](https://doi.org/10.1038/s41597-022-01432-0).

- LIST Johann-Mattis *et al.*, 2022, “Lexibank Analysed”, *Scientific Data* vol. 316, n° 9, Zenodo, 17 avril 2023, DOI : 10.5281/zenodo.7836668.
- MACNEILAGE Peter F. & DAVIS Barbara L., “On the origin of internal structure of word forms”, *Science*, vol. 288, n° 5465, p. 527-31, DOI : 10.1126/science.288.5465.527.
- MORAN P. A. P., 1950, “Notes on Continuous Stochastic Phenomena”, *Biometrika*, vol. 37, n° 1/2, p. 17-23, DOI : 10.2307/2332142.
- MURDOCK George Peter, 1959, “Cross-Language Parallels in Parental Kin Terms”, *Anthropological Linguistics*, vol. 1, n° 9, p. 1-5.
- NICHOLS Johanna, 1992, *Linguistic diversity in space and time*, Chicago, University of Chicago press.
- NICHOLS Johanna, 1994, “The spread of language around the Pacific rim”, *Evolutionary Anthropology: Issues, News, and Reviews*, vol. 3, n° 6, p. 206–15.
- PASSMORE Sam, BARTH Wolfgang, GREENHILL Simon J., QUINN Kyla, SHEARD Catherine, ARGYRIOU Paraskevi, BIRCHALL Joshua, *et al.*, 2023, “Kinbank: a global database of kinship terminology”, *PLOS ONE* vol. 18, n° 5, DOI : 10.1371/journal.pone.0283218.
- RINGE Donald A., 1999, “How hard is it to match CVC-roots?”, *Transactions of the Philological Society*, vol. 97, n° 2, p. 213-44, DOI : 10.1111/1467-968X.00049.
- RINGE Donald A., 1992, “On Calculating the Factor of Chance in Language Comparison”, *Transactions of the American Philosophical Society*, vol. 82, n° 1, p. 1-110, DOI : 10.2307/1006563.
- RINGE Donald A., 1996, “The Mathematics of ‘Amerind’”, *Diachronica*, vol. 13, n° 1, p. 135-54, DOI : 10.1075/dia.13.1.06rin.
- RUHLEN Merritt, 2007, *L’origine des langues : sur les traces de la langue mère*, trad. BANCEL Pierre, Gallimard, Paris.
- RZYMSKI Christoph, TRESOLDI Tiago, GREENHILL Simon J., WU Mei-Shin, SCHWEIKHARD Nathanael E., KOPTJEVSKAJA-TAMM Maria, GAST Volker, *et al.*, 2020, “The Database of Cross-Linguistic Colexifications, Reproducible Analysis of Cross-Linguistic Polysemies”, *Scientific Data*, vol. 7, n° 1, DOI : 10.1038/s41597-019-0341-x.
- SAPIR Edward, 1916, *Time perspective in aboriginal American culture: a study in method*, Government Printing Bureau, Ottawa.

SCHEER Tobias & SÉGÉRAL Philippe, 2001, « La coda-miroir », *Bulletin de la Société de Linguistique de Paris*, vol. 96, n° 1, p. 107-52.

TRASK Robert Lawrence, 1996, *Historical linguistics*, Arnold, Londres.

URBAN Matthias, 2012, “Analyzability and semantic associations in referring expressions: a study in comparative lexicology”, Leiden University, URL : <https://hdl.handle.net/1887/19940>.

WICHMANN Søren, HOLMAN Eric W. & BROWN Cecil H., 2022, “CLDF dataset derived from Wichmann *et al.*, ‘ASJP Database’ v20 from 2022”, Zenodo, DOI : 10.5281/zenodo.7079637.

WICHMANN Søren, HOLMAN Eric W. & BROWN Cecil H., 2010, “Sound symbolism in basic vocabulary”, *Entropy*, vol. 12, n° 4, DOI : 10.3390/e12040844.

Résumé : Cet article teste l’hypothèse selon laquelle certains patrons son/sens associés à « mère » et « sein » pourraient refléter une codiffusion très ancienne, héritée des premières migrations d’*Homo sapiens* et de la centralité de la dyade mère-enfant. Dans une perspective d’« archéologie linguistique », quatre traits phonotactiques sont examinés dans 2 959 langues (« mère ») et 7 322 langues (« sein ») issues des bases Lexibank et ASJP : [n]/[ŋ] et [na]/[ŋa] en initiale de « mère » ; [mu] et [amu] pour « sein ». Leur distribution est évaluée via des analyses spatiales (Moran I, z-scores sur grilles 2°×2°, permutations aléatoires, distances orthodromiques). Les résultats révèlent une structuration non aléatoire pour [n]/[ŋ] et [mu], avec des foyers en Afrique, Asie du Sud, Insulinde, Mélanésie et Australie, soit des zones concordant avec les routes australes des premières dispersions d’*homo sapiens*. Les formes [na] et [amu], plus localisées, apparaissent comme des archaïsmes régionaux. La forte cooccurrence géographique entre [n]/[na] et [mu], contrastant avec la diffusion restreinte de [amu], suggère un noyau lexical ancien liant « mère » et « sein », non réductible aux seuls biais articulaires. Sans postuler de protolangage, l’étude montre que l’aréologie linguistique permet d’identifier des traces fossiles d’une structuration culturelle précoce autour de la dyade mère-enfant, offrant un appui partiel à l’hypothèse d’une unité culturelle initiale diffusée avec les premières sorties d’Afrique d’*Homo sapiens*.

Mots clés : archéologie linguistique, dyade mère-enfant, phonotactique, diffusion lexicale, routes migratoires humaines, aréologie

When Words Tell Things: A Linguistic Archaeology of the Mother–Child Dyad

Abstract: *This paper tests the hypothesis that certain sound–meaning patterns associated with “mother” and “breast” may reflect a very ancient codiffusion, inherited from the earliest migrations of Homo sapiens and from the centrality of the mother–infant dyad. Within a framework of “linguistic archaeology,” four phonotactic traits are examined across 2,959 languages (“mother”) and 7,322 languages (“breast”) from the Lexibank and ASJP databases: [n]/[ŋ] and [na]/[ɲa] in initial position for “mother,” and [mu] and [amu] for “breast.” Their distribution is assessed through spatial analyses (Moran’s I, binomial z-scores on 2°×2° grids, random permutations, great-circle distances). The results reveal a non-random structuring for [n]/[ŋ] and [mu], with hotspots in Africa, South Asia, Island Southeast Asia, Melanesia, and Australia, regions that align with the southern routes of early Homo sapiens dispersal. The forms [na] and [amu], more geographically restricted, appear as regional archaisms. The strong geographic co-occurrence between [n]/[na] and [mu], contrasting with the limited spread of [amu], suggests an ancient lexical core linking “mother” and “breast,” not reducible to articulatory biases alone. Without positing a single protolanguage, the study shows that linguistic areology can reveal fossil traces of an early cultural structuring around the mother–infant dyad, providing partial support for the hypothesis of an initial cultural unity disseminated during the first out-of-Africa dispersals of Homo sapiens.*

Keywords: *linguistic archaeology, mother–infant dyad, phonotactics, lexical diffusion, human migratory routes, areology*

