

# Présentation générale

Konstantin Pozdniakov\*

## TENDANCES ACTUELLES EN COMPARATISME ET RECONSTRUCTION

Les publications récentes en linguistique historique et comparée sont si nombreuses que même leur survol complet n'est pas envisageable dans le format de cette présentation : des dizaines de monographies divulguent les résultats concrets d'études comparatives dans des domaines encore peu ou pas étudiés jusqu'à récemment. Dans le même temps, des manuels ou traités théoriques continuent d'être consacrés à ce domaine. Comme par le passé, on regrette le manque d'ouvrages de vulgarisation de qualité susceptibles de faire connaître à un plus large public les enjeux de la linguistique comparée, bien que quelques titres récents méritent l'attention<sup>1</sup>. Plus encore, on manque de monographies où les auteurs s'appuieraient sur des avancées pratiques nouvelles pour alimenter les débats anciens mais toujours d'actualité sur la théorie et la méthodologie de la linguistique comparée. C'est dans cette niche étroite, mais qui à notre sens correspond à un besoin réel que le présent volume espère s'inscrire.

Tous les articles de ce volume constituent le développement de communications présentées au séminaire «Linguistique comparative historique au XXIème siècle : enjeux théoriques et méthodologiques», que nous avons créé à l'INALCO dans le cadre de notre chaire de l'Institut universitaire de France en 2011. Le cahier des charges proposé aux invités était le suivant : a) présenter des résultats pratiques récents et originaux obtenus dans leur domaine de spécialité, b) procéder à une mise en perspective théorique de ces résultats, identifier et présenter sous leur éclairage au moins une des questions théoriques et/ou méthodologiques actuelles de la linguistique historique et comparée. Cette approche, allant des résultats concrets aux questions théoriques et méthodologiques, constitue le fil conducteur essentiel de ce volume où ces communications sont développées par leurs auteurs.

Pour faciliter l'accès des non spécialistes aux débats théoriques évoqués ici, il est nécessaire de formuler brièvement quels postulats de la linguistique comparée sont aujourd'hui considérés comme établis, et quelles positions théoriques sont

---

\* INALCO, IUF, LLACAN. Courriel : konstantin.pozdniakov@cns.fr. Ce travail est lié au programme «Investissements d'Avenir» géré par l'Agence Nationale de la Recherche ANR-10-LABX-0083 (Labex EFL).

<sup>1</sup> Il convient de mentionner ici la dernière publication d'un des auteurs de ce volume, G. Starostin, 2015, Г. Старостин. *К истокам языкового разнообразия. [Aux sources de la diversité linguistique]*, Moscou, Editions Delo.

encore (ou devraient être) l'objet de débats. Pour cela, il est utile de procéder à une distinction entre la **méthode** historique et comparée d'une part, et la **problématique** historique et comparée de l'autre. De fait, la méthode historique et comparée, qui jouit à juste titre de la réputation d'être la plus rigoureuse des sciences humaines, ne constitue qu'une petite partie de l'étude de l'histoire des langues apparentées, sachant qu'une couche considérable des problèmes de la linguistique comparée dépasse le cadre de la méthode historique et comparée dans son acception traditionnelle. Plus encore, la partie de la problématique qui relève de la méthode historique et comparée fait l'objet d'un certain consensus (les postulats théoriques ne sont pas remis en questions et sont acceptés par la quasi-totalité des spécialistes), alors que l'autre partie, qui n'est pas directement en prise avec ces postulats, fait l'objet d'âpres débats, dans lesquelles s'exprime une très large palette d'avis de spécialistes, et qui sont toujours ouverts.

Nous commencerons par une présentation du champ de la méthode historique et comparée, puis examinerons les problèmes centraux de la linguistique comparée aujourd'hui, qui sont privés de toute possibilité de s'appuyer sur cette méthode. Ce qui suit ne prétend pas à l'exhaustivité, et se limite aux aspects de la question abordés dans les articles du volume.

## 1. LA MÉTHODE HISTORIQUE ET COMPARÉE

On lit souvent dans les manuels que la méthode historique et comparée est fondée sur les correspondances phonétiques régulières observées entre langues apparentées. Certes, mais les correspondances phonétiques ne sont que la conséquence du principe fondamental de la méthode historique et comparée, à savoir que les changements phonétiques diachroniques doivent être **systemiques**. Si la séquence latine **\*kt** change indépendamment dans plusieurs langues-filles, cela doit concerner tous les mots correspondant à un contexte déterminé : **tt** en italien, **pt** en roumain, **Vt** en français et en portugais, **cc** en espagnol etc.

Une fois les correspondances mises en évidence, il faut les vérifier. Ainsi, sans connaître le roumain, on peut calculer la forme du mot signifiant 'poitrine' dans cette langue, à supposer que le réflexe de la protolangue n'ait pas été évincé par un mot emprunté. En passant par le portugais *peito* et l'espagnol *pecho*, nous aboutissons à la forme roumaine *piept-*, puis nous consultons un dictionnaire du roumain... qui donne la forme *piept*. De la même manière, à partir du français *lait*, *nuit*, *fait*, *huit* nous pouvons retrouver les mots roumains *lapte*, *noapte*, *fapt*, *opt*. Cette possibilité de vérification expérimentale est une des plus extraordinaires disponibles dans les sciences humaines.

Un des résultats manifestes de cette méthode est de permettre de démontrer que des langues sont apparentées en mettant à jour des correspondances régulières. Il s'agit finalement d'un argument statistique. La probabilité que la régularité soit la conséquence d'une origine commune des langues est bien supérieure à celle du hasard.

La régularité des correspondances n'est pas le seul «bonus» que les langues offrent aux comparatistes. Il existe au moins deux autres propriétés qui ne peuvent relever du hasard et qui permettent d'identifier avec certitude une

parenté entre les formes et les langues comparées. D'une part, le plus souvent, bien que cela ne soit pas obligatoire, les correspondances entre langues sont non seulement régulières, mais aussi proportionnelles. L'article d'Alexandre François en fournit une illustration, en présentant les correspondances entre 17 langues du Vanuatu Nord. Nous voyons notamment que dans la langue lemerig, le proto-océanique \*p donne régulièrement v, en d'autres termes, l'occlusive labiale sourde devient une fricative sonore. Mais parallèlement, nous constatons que ce changement régulier n'est pas isolé, puisqu'on observe également que \*k > ɣ, ce qui révèle un changement proportionnel des sourdes aux sonores et d'occlusives à fricatives. La découverte de ce type de proportions diachroniques constitue un argument supplémentaire de poids pour établir la parenté des langues.

Si l'on cherche bien, on peut trouver encore quelques petits «bonus», par exemple, le bonus d'un «principe de domino» qui se manifeste assez fréquemment. On observe dans des langues assez variées que des lacunes apparues dans les systèmes phonologiques (du fait de la transformation d'un son en un autre) se trouvent comblées grâce à de nouveaux changements. Le même tableau des correspondances Vanuatu Nord montre qu'en lemerig, après le changement \*k > ɣ, c'est-à-dire après la disparition du \*k, un nouveau changement \*ɣ > k est intervenu<sup>2</sup>. C'est sur ce type de changements qu'est bâtie la loi d'anthologie de Grimm qui rend compte des changements intervenus du proto-indo-européen au proto-germanique par une chaîne de transformations comme \*k > h, \*g > k, \*gh > g. Ce type d'enchaînement n'a rien de mystérieux. La disparition du k fait perdre au trait de voisement sa fonction distinctive, et ainsi g passe à k. Par enchaînement, le changement du g prive de sa fonction le trait supplémentaire (l'aspiration) de gh, qui disparaît également. Parmi ce type de changements en chaîne, le plus long est à ma connaissance celui qui est observé en basari et en bedik (branche Nord de la famille atlantique) avec l'enchaînement des 5 changements suivants: \*s > š, r > s, \*l > r, \*d > l, \*n > d.<sup>3</sup>

Ainsi, la découverte de correspondances régulières, mais encore proportionnelles, et a fortiori avec des processus de substitution successive, constitue un bel acquis scientifique, apportant la preuve irréfutable d'une parenté entre les langues comparées. Mais malheureusement, cela s'arrête là. Pour aller plus loin dans la connaissance historique des langues concernées il faut chercher de nouveaux cadres méthodologiques. La méthode historique et comparée classique est donc extrêmement efficace, mais son application reste relativement limitée. Elle est démunie face au défi de la reconstruction sémantique, pourtant incontournable si on souhaite que la linguistique comparée, par-delà la reconstruction phonétique, permette d'acquérir des connaissances sur les proto-cultures. De plus, sans reconstruction sémantique fiable il n'est pas possible de

<sup>2</sup> L'ordre inverse (\*ɣ > k suivi par \*k > ɣ) est exclu puisque cela aurait conduit à la disparition de k (<\*ɣ), qui se serait transformé en ɣ, comme \*k (\*ɣ > k > ɣ).

<sup>3</sup> K. Pozdniakov, 2015, Diachronie des classes nominales atlantiques : morphophonologie, morphologie, sémantique, in D. Creissels et K. Pozdniakov (éds.), *Les classes nominales dans les langues atlantiques*, Köln, Rüdiger Köppe Verlag, p. 57-102.

comparer entre elles les protolangues issues de la reconstruction, afin d'aboutir à reconstruire les protolangues plus anciennes. Par ailleurs, la méthode historique et comparée ne fournit pas d'éléments décisifs permettant de reconstruire la forme phonétique des mots: même les néo-grammairiens, pourtant férus de méthode historique et comparée, se sont abstenus de reconstruire les mots de la protolangue. Cette méthode ne permet pas non plus d'interpréter les changements phonétiques irréguliers, notamment les changements par analogie. On ne peut s'appuyer sur elle pour reconstruire la syntaxe. Elle ne fournit pas d'arguments fiables pour bâtir une classification généalogique des langues. Elle est impuissante face à la question de savoir qui parlait ces protolangues reconstruites, quand, et où. Toutes ces facettes essentielles de la problématique comparatiste, ne peuvent pas être traitées par la méthode historique et comparée.

Il est important de souligner cela ici pour deux raisons.

1) La confusion entre la vaste sphère de la problématique du comparatisme d'une part, et le domaine assez restreint mais extrêmement précis de la méthode destinée à répondre concrètement à la question de la parenté des langues et unités linguistiques comparables, est source des nombreux malentendus qui émergent régulièrement aussi bien dans les textes de vulgarisation que dans les pages de revues scientifiques prestigieuses. Par exemple, la critique du modèle le plus courant de l'arbre généalogique (en soi tout à fait légitime), conduit à des remises en cause de la méthode comparatiste classique dans son ensemble, ce qui est beaucoup moins légitime puisque la méthode historique et comparée ne présuppose en elle-même aucun modèle de divergence des langues, même si elle pousse le chercheur à recourir au modèle d'un graphe simple, du fait que la régularité des correspondances implique une période de développement indépendant des langues-filles.

2) Dissocier la problématique et la méthode est essentiel pour comprendre les contours de la polémique théorique actuelle en comparatisme, ce qui à son tour permet de mieux apprécier l'apport des auteurs de ce volume pour le domaine. L'existence de correspondances entre langues apparentées est aujourd'hui un acquis pratiquement incontestable, accepté comme une donnée par toutes les écoles théoriques. Au contraire, tous les problèmes qui sortent du champ de la méthode historique et comparée, et dont beaucoup ont été mentionnés plus haut, restent ouverts aux débats théoriques, suscitant parfois même jusqu'à aujourd'hui des points de vue diamétralement opposés. C'est dans ce contexte que s'inscrivent les débats abordés par les auteurs de ce volume, ce qui explique le sous-titre 'tendances actuelles'.

Pour situer les idées discutées dans ce volume dans un contexte théorique précis, revenons à certaines questions essentielles de la problématique du comparatisme, qui se trouvent comme par le passé sur le fil du rasoir de débats contradictoires.

## 2. CORRESPONDANCES IRRÉGULIÈRES ET CHANGEMENTS PAR ANALOGIE

Les changements diachroniques irréguliers, pour lesquels la méthode historique et comparée n'est pas efficace, jouent un rôle particulièrement

important en grammaire, ainsi qu'au sein des classes lexicales sémantiquement les plus structurées comme le système des numéraux, des termes de parenté, des termes d'orientation spatiale, ou encore de mesure ou autres. D'après Saussure : «... l'histoire de chaque langue permet de découvrir un fourmillement de faits analogiques accumulés les uns sur les autres, et, pris en bloc, ces continus remaniements jouent dans l'évaluation de la langue un rôle considérable, plus considérable même que celui des changements de sons. (...) La langue est une robe couverte de rapiécages faits avec sa propre étoffe»<sup>4</sup>. La raison principale de l'irrégularité de ces changements est que les affixes grammaticaux et groupes de mots concernés sont organisés en paradigmes, au sein desquels sont à l'œuvre des lois de régularisation par analogie de segments de paradigmes encore mal étudiées, et qui impliquent une correspondance totale ou partielle des formes. C'est-à-dire que contrairement aux changements phonétiques réguliers, qui sont plutôt syntagmatiques (les sons changent dans un contexte phonotactique particulier), les changements par analogie sont paradigmatiques.

Examinons par exemple les changements dans le paradigme de la mesure spatiale en russe. Dans le Tableau 1 on a huit formes d'adjectifs comparatifs, toutes «incorrectes» du point de vue de la diachronie.

Tableau 1

	Formes «incorrectes» mais existantes	Formes «correctes» mais inexistantes
«plus»		
'plus haut'	výš-e	*vys-óč-e
'plus profond'	glúbž-e	*glub-óč-e
'plus large'	šír'-e	*šir-óč-e
'plus loin'	dál'š-e	*dal'-éč-e
«moins»		
'plus bas'	níž-e	*niz-č-e
'moins profond'	mél'č-e	*mél-č-e
'plus étroit'	úz-e	*úz-č-e
'plus proche'	blíž-e	*blíz-č-e

Le passage de formes trisyllabiques aux formes disyllabiques, ainsi que des consonnes dures aux consonnes palatalisées, a érigé le trait «disyllabique +» et «la palatalisation +» en traits submorphiques par lesquels ce paradigme se distingue des paradigmes voisins, comme on le voit dans les tables 2 à 4.

Dans le paradigme de base (Tableau 2), à la différence de ce qui est observé dans les formes comparatives, les adjectifs représentant les pôles + et - se distinguent aussi bien par leur nombre de syllabe que par la place de l'accent tonique.

<sup>4</sup> de Saussure F., 1972, *Cours de linguistique générale*, Paris, Payot, p. 235.

Tableau 2

«plus»		«moins»	
'haut'	vys-ók-ij	'bas'	níz-k-ij
'profond'	glub-ók-ij	'pas profond'	mél-k-ij
'large'	šir-ok-ij	'étroit'	úz-k-ij
'loin'	dal'-ók-ij	'proche'	blíz-k-ij

On observe le même principe dans le paradigme des adverbes (Tableau 3), à la seule différence que l'accent dans les adjectifs de polarité positive est final. Notons que parmi 19337 adverbes à base polysyllabiques, ces 4 adverbes sont les seuls ayant l'accent sur la syllabe finale.

Tableau 3

«plus»		«moins»	
'haut'	vys-ok-ó	'bas'	níz-k-o
'profond'	glub-ok-ó	'pas profond'	mél-k-o
'large'	šir-ok-ó	'étroit'	úz-k-o
'loin'	dal'-ek-ó	'proche'	blíz-k-o

Enfin, dans le paradigme des formes brèves des adjectifs (Tableau 4), les adjectifs de polarité positive vs négative s'opposent par la place de l'accent, et non par le nombre de syllabes :

Tableau 4

«plus»		«moins»	
'haut'	vys-ók	'bas'	níz-ok
'profond'	glub-ók	'pas profond'	mél-ok
'large'	šir-ók	'étroit'	úz-ok
'loin'	dal'-ók	'proche'	blíz-ok

On a ici une concentration de formes historiquement «incorrectes» particulièrement élevée. Ces tableaux montrent que les changements par analogie n'opèrent pas uniquement au sein d'un même paradigme. Les changements irréguliers permettent de créer un système complexe d'oppositions et de neutralisations au sein d'un groupe de paradigmes proches, quant aux formes elles-mêmes, elles acquièrent un caractère presque iconique.<sup>5</sup>

Dans ce recueil c'est l'article de P. Kocharov qui présente un grand nombre de formes analogiques dans le domaine de la morphologie du verbe indo-européen. L'auteur remarque notamment que «*The limitations of lexical reconstruction and the work of morphological analogy are the two major hindrances in the research*

<sup>5</sup> L'iconicité de ces huit adjectifs a été remarquée pour la première fois par André Zalizniak (A. A. Zalizniak, 2002, *Русское именное словоизменение* [La flexion nominale russe], Moscou, JaSK, c. 534-542.

of the PIE aspect». De fait, pour le comparatiste, les changements par analogie représentent un obstacle considérable et souvent infranchissable à la reconstruction. En accédant à la régularité des formes au sein des paradigmes, la langue sacrifie la régularité les changements historiques, et de ce fait, les morphèmes grammaticaux, qui dans un certain nombre de cas sont plus stables que le lexique, se révèlent dans d'autres cas moins appropriés pour la reconstruction que les racines lexicales.

C'est pour cette raison qu'il est souvent difficile de reconstruire les systèmes pronominaux, dans lesquels les processus analogiques, appliqués à des formes généralement très courtes, se déroulent de manière particulièrement active. Cela est illustré par trois paradigmes pronominaux de langues de familles différentes (Tableau 5) :

Tableau 5

	Français		Erzya		Dogon	
	SG	PL	SG	PL	SG	PL
1	mon	més	mon	myn'	mù	èmé
2	ton	tes	ton	tyn'	ù	é
3	son	ses	son	syn'	vò	bé

Dans ces paradigmes chaque son de chacune des formes «œuvre» pour l'alignement analogique, devenant le marqueur submorphique d'une personne ou d'un nombre donné (en erzya et en dogon, il ne s'agit pas seulement d'un son, mais d'un trait distinctif de phonème : en erzya la palatalisation marque le pluriel, en dogon, la labiale nasale marque la première personne, et la labiale orale la troisième personne ; en dogon, le ton et la structure V marque la deuxième personne).

L'article d'Alexandre François illustre ici particulièrement bien la complexité de la reconstruction des pronoms. Au passage, comme l'a d'ailleurs signalé un des rapporteurs, dans ce genre d'étude l'analyse des alignements analogiques est incontournable. Quand l'auteur conclut que «*Among the 41 <...> isoglosses, only 24 are compatible with the tree <...> : the other 17 are not, as they straddle across branches*», on voit que si l'on écarte les cas de changement analogique qui peuvent se manifester parallèlement dans différentes langues, le pourcentage de formes qui semblent mettre en question le modèle arborescent sera bien moindre.

Le changement par analogie est également important pour les problématiques soulevées par Guillaume Jacques et Victor Porkhomovsky.

### 3. LA PROTOLANGUE ET SA RECONSTRUCTION. TYPOLOGIE ET COMPARATIVISME

Comme nous l'avons vu, la méthode historique et comparée dans son acception classique ne fournit pas de réponses directes sur les protolangues elles-

mêmes. Ainsi, en caricaturant les affirmations publiées sur les protolangues, on pourrait les ramener à deux positions principales : 1) la protolangue a existé, 2) la protolangue n'a pas existé. La dernière réponse suppose que nos reconstructions ne sont qu'un «code algébrique» commode, mais artificiel, pour interpréter des correspondances phonétiques. Bien entendu, la deuxième position, qui circule de manuel en manuel, provient d'une lecture naïve (et tout simplement erronée) des néo-grammairiens ou d'Antoine Meillet. Car il n'y a guère de différence entre le fait de nier l'existence de la protolangue et celui de nier qu'il soit possible de la reconstruire de manière exacte. Or, c'est bien ce dernier point qui occupait les classiques de la méthode historique et comparée, ce qui ne les a pas empêchés d'énoncer une série de jugements fondamentaux sur les protolangues, qui attendent encore d'être discutés en profondeur.

Que s'agit-il de reconstruire ? Une des réponses possibles, qui me semble-t-il pourrait être acceptée par de nombreux praticiens du comparatisme, est que nous reconstruisons une protolangue réelle que personne n'a jamais parlée. Le paradoxe est assez simple : l'individu humain ne vit pas cinq siècles ; il ne vit pas non plus normalement dans 2 villages distants de 300 km, alors que les éléments de protolangue que nous parvenons à reconstruire ne sont vraisemblablement pas synchronisés dans le temps ni dans l'espace. Mais cela ne signifie pas que les hypothèses des comparatistes concernant la reconstruction ne s'appuient sur rien. Par exemple, pour mieux synchroniser les reconstructions lexicales, on peut essayer de «remonter d'un niveau», c'est-à-dire de passer de la reconstruction du lexique d'un groupe à la reconstruction du lexique d'une famille, puis revenir en arrière afin de tenter d'affiner la variation dialectale de la protolangue.

Pour vérifier leurs reconstructions les comparatistes utilisent aujourd'hui des données très diverses. La question de l'usage qui peut être fait des données de la typologie linguistique est très présente dans les débats théoriques sur la reconstruction des protolangues. La recevabilité des arguments typologiques pour vérifier des reconstructions s'appuie sur le fait que les protolangues semblent fonctionner de manière comparable aux langues modernes. A titre personnel, je peux dire que pour moi la découverte la plus importante du 20<sup>ème</sup> siècle en comparatisme est que dans la langue il n'y a pas de progrès. Les langues ne deviennent ni plus complexes, ni plus performantes, elles se contentent de se transformer. Les reconstructions ayant la plus grande profondeur chronologique (par exemple nostratique ou afro-asiatique) montrent dans les langues les plus anciennes les mêmes propriétés que celles qui sont observables aujourd'hui dans les langues modernes. De ce fait la typologie et notamment la typologie diachronique se révèle être un appui puissant à la reconstruction, notamment grammaticale. Dans ce volume, l'efficacité du recours à la typologie pour la reconstruction grammaticale est abordée par les articles de Guillaume Jacques et Victor Porkhomovsky.

Dans la typologie diachronique des langues apparentées, il n'est pas toujours aisé de comprendre si les données typologiques confirment la reconstruction ou si elles l'infirmement. Ainsi dans le cas d'école des trois séries de consonnes reconstruites pour l'indo-européen, qui incluent une opposition d'aspiration dans les séries de sonores (pas représentée parmi les sourdes), il semble que les



données de la typologie des langues indo-européennes ne soutiennent pas la reconstruction, puisque ce type de système n'est attesté dans aucune langue indo-européenne. Mais les données de la typologie diachronique, notamment des langues indo-européennes, montrent qu'une situation où un élément de la protolangue n'est représenté dans aucune langue-fille est tout à fait possible. Ainsi, si on s'appuie sur l'exemple cité au début de cet article, on peut constater qu'aucune des langues (et, comme disent les romanistes, semble-t-il, aucun des dialectes romans) n'a conservé \*kt, qui n'a survécu que dans des emprunts au latin comme les mots français *nocturne*, *octave*, *lactaire*, *facture*, qui cohabitent avec les réflexes réguliers fr. *nuit*, *huit*, *lait*, *fait*<sup>6</sup>. A cet égard, à la lumière de la typologie diachronique, il n'y a rien de séditieux dans le fait que Valentin Vydrin, dans ce volume, parvienne à la conclusion que «*The reconstruction of the tonal split is a strong argument for the hypothesis of a 2-level system in the Proto-Southern Mande. However, there is not a single modern language in this group that has retained such a system; most of them have 3 level tones, some others (Dan-Blo, Kla-Dan, Tura, Goo, Gban, Yaure) have 4 level tones, and Dan-Gweetaa has 5 level tones (apart from the contour tones)*».

Remarquons que dans le cas de langues apparentées il n'est pas toujours facile de déterminer si un trait partagé est imputable à la typologie ou à la généalogie. J'en donnerai un exemple significatif. De nombreux travaux en linguistique africaine considèrent que l'existence même de classes nominales dans une langue ne peut pas être un critère permettant de rattacher la langue au groupe Niger-Congo, du fait que ce trait est considéré comme typologique et non généalogique. Après avoir conduit des recherches sur ces classes Niger-Congo pendant 40 ans, je peux affirmer en toute conscience que quelle que soit la diversité des systèmes de classes nominales, surtout dans les langues d'Afrique de l'Ouest, il existe une série de traits vérifiés qui permettent d'affirmer immédiatement que le système des classes nominales, par exemple en limba (dont la place dans la classification généalogique n'est pas totalement établie), relève résolument du Niger-Congo, ce qui n'est pas le cas, par exemple, du tchéchéne.

Pourquoi dans les reconstructions les protolangues apparaissent-elles souvent typologiquement différentes des langues-filles ? Le plus souvent cela est le revers de la médaille (parfois inévitable) de la méthode de reconstruction, comme le montrent deux cas typiques.

Premier cas : En établissant des correspondances régulières dans une famille de langues, il est nécessaire, notamment à l'étape initiale, de ramener chaque série de correspondances à un proto-phonème donné. Ainsi, en examinant les trois séries de correspondances régulières du Tableau 6 entre labiales non initiales dans deux langues étroitement apparentées de la famille atlantique, le basari et le konyagi, nous sommes amenés à distinguer trois proto-phonèmes (désignés ici \*1, \*2, \*3 sans tenir compte de leur reconstruction).

<sup>6</sup> Cela ne signifie-t-il pas que le latin *kt*, était déjà condamné en synchronie, et que les langues-filles héritent donc non seulement des éléments de la protolangue, mais aussi de ses conflits internes, qu'elles résolvent plus tard comme ils peuvent ?

Tableau 6

Glose	Basari	Konyagi	Glose	Basari	Konyagi	Glose	Basari	Konyagi
	*1			*2			*3	
	mb	mp		m	m		m	mb
changer	l̩mbátá	n̩mpát	père	s̩m	r̩m	chanvre	θámá	bámá
épine	d̩mbán	l̩mpáĩ	sans	k̩m	k̩m	hanche	p̩m / f̩m	f̩mbár / p̩mbár
tordre	s̩mb	t̩mp / r̩mp	froid	d̩m	f̩m	pousser	f̩m	p̩mbá

Cependant dans ces exemples l'attention est attirée sur les points suivants : dans la série \*3, les consonnes initiales sont également labiales, alors que dans la série \*2 elles ne le sont pas. On peut montrer que cette répartition complémentaire est apparue comme conséquence d'une réalisation différente du même proto-phonème dans des conditions différentes : en konyagi \*mb > mp ; \*m > m après non-labiales / > mb après labiales<sup>7</sup>. Dans ce cas on reconstruit deux phonèmes et non trois. En réalité, surtout à l'étape initiale de la recherche, alors que ces règles n'ont pas encore été mises en évidence, il y a beaucoup plus de proto-phonèmes reconstruits que dans n'importe laquelle des langues-filles. Au fur et à mesure des progrès de la recherche cette divergence typologique disparaît.

Le deuxième cas d'«anomalie typologique» est lié aux tentatives d'imputer à la protolange des propriétés de «pré-langue», de tenter d'expliquer la formation des mots et des catégories grammaticales de langues apparentées «à partir de zéro», depuis la création du Monde. Pour en trouver une illustration nous pouvons recourir cette fois encore aux classes nominales Niger-Congo. Dans une série de travaux bien connus, on tente de formuler les étapes initiales de la formation des classes nominales en proto-niger-congo et même dans sa langue-fille la plus connue, le proto-bantu. Pourtant il est tout à fait clair qu'en proto-bantu, le système des classes était non seulement déjà constitué, mais il avait selon toute vraisemblance déjà subi des changements substantiels par rapport aux états précédents. On a toutes les raisons de penser qu'il existait déjà un système de classes développé en proto-niger-congo, même s'il a été perdu par la suite dans une série de langues, comme par exemple les langues mandé. Les tentatives d'expliquer les classes «depuis Adam» font étonnamment bon ménage avec la tendance à représenter les classes de la protolange comme un système plus cohérent, qui aurait perdu sa rigueur dans toutes les langues-filles. Cette approche est particulièrement manifeste dans les descriptions de la sémantique des classes nominales. Ainsi, chez de nombreux auteurs, l'analyse de la classe conventionnellement nommée «classe des animaux» s'accompagne de deux remarques fréquentes : 1) «dans la langue moderne les noms d'animaux ne font

<sup>7</sup> C'est une nouvelle illustration du principe de substitution de niche mentionné plus haut.

plus tous partie de cette classe», et 2) «dans la langue moderne cette classe n'est plus strictement limitée aux noms d'animaux». Pourtant, il est tout à fait évident que parmi les centaines de langues ayant des classes nominales, aucune ne présente une classe comportant **tous** les noms d'animaux et **uniquement** des noms d'animaux. Pourquoi en serait-il autrement dans la protolange ?

Il est clair que nos connaissances sur des protolanges distantes de plusieurs millénaires restent très limitées. Même la connaissance de l'indo-européen, dont la reconstruction des morphèmes est pratiquement achevée, ne cesse de progresser. C'est ce que traduit dans ce volume l'article particulièrement significatif de Romain Garnier, qui formule ainsi ce problème: «*Mon propos n'est pas de nier la réalité linguistique de l'indo-européen, mais de décharger la langue commune d'un certain nombre de fantômes, qui peuplent encore jusqu'aux plus récents dictionnaires étymologiques, et qui ont passé en doctrine*».

#### 4. LE LEXIQUE DE «BASE» ET LE LEXIQUE «CULTUREL». PROBLÈMES DE LA RECONSTRUCTION SÉMANTIQUE

En procédant à la reconstruction des protolanges, nous espérons également parvenir à reconstruire des éléments des cultures anciennes. Cependant, en matière d'histoire linguistique préhistorique, la reconstruction d'éléments culturels demande en première instance de s'écarter du lexique spécifiquement culturel pour s'atteler à la reconstruction de termes les plus neutres possibles par rapport au domaine, et notamment le lexique de base, le moins susceptible d'être emprunté, et donc le plus apte à la reconstruction.

Le développement du comparatisme au-delà du domaine indo-européen a énormément bénéficié des idées exprimées dans les années 1960 par Maurice Swadesh. Cependant, Valentin Vydrin remarque très justement dans ce volume : «*It is true that since its invention in mid-20 century, glottochronology and lexicostatistics have been considerably discredited because of the numerous shortcomings of M. Swadesh's method, to the point that many comparative linguists have grown allergic to the very mention of glottochronology.*»

Sans nous arrêter ici en détail sur la lexicostatistique et la glottochronologie, auxquelles sont consacrées des dizaines de monographies théoriques et des centaines de travaux empiriques, nous voulons en souligner quelques aspects qui sont source de très nombreux malentendus, y compris dans les débats entre comparatistes professionnels.

Premièrement, comme tout un ensemble d'autres problèmes examinés ici, la lexicostatistique et la glottochronologie n'ont pas de rapport direct avec les procédures de la méthode historique et comparée.

Deuxièmement, un certain nombre de postulats fondamentaux sur lesquels sont construites les méthodes de lexicostatistique devraient pouvoir faire consensus, comme: a) plus le temps passe et moins la langue conserve lexique hérité ; b) le lexique de base est moins soumis aux emprunts que le lexique culturel, par conséquent, l'éviction du vocabulaire de base est moins dépendante des contacts et des cataclysmes culturels, ce qui à son tour signifie que la vitesse de

disparition du vocabulaire de base est plus universelle et plus susceptible d'être analysée et calculée.

Troisièmement, depuis l'époque des premières publications de Swadesh, la lexicostatistique et la glottochronologie ont connu des changements majeurs, en premier lieu grâce aux travaux théoriques de Serge Starostin. La formule et les méthodes de calcul ont changé, la liste de 100 mots s'est réduite de moitié, notre vision concernant la vitesse d'extinction du vocabulaire de base a beaucoup évolué, les résultats empiriques considérables accumulés ont notamment montré qu'au sein de la liste de Swadesh, les termes n'ont pas tous le même degré de stabilité. Cependant, aujourd'hui encore on entend et on lit souvent des critiques adressées à la lexicostatistique, en décalage complet avec les résultats qu'elle a permis d'obtenir dans les dernières décennies, et que les linguistes ignorent le plus souvent. Un exemple : les spécialistes de lexicostatistique sont souvent critiqués pour une attention insuffisante aux reconstructions sémantiques, et notamment à la typologie des changements sémantiques. Martin Haspelmath note : «*Some proposals have been made for identifying the most stable lexical meanings (e.g. Dyen et al. 1967, Lohr 1999), but none of these has been based on a broad database from a representative sample of the world's languages*»<sup>8</sup>.

Notons qu'en 2004, dans son article fondamental «Définir la stabilité du lexique de base» [Starostin 2007]<sup>9</sup>, Serge Starostin a déterminé, sur la base d'un matériau empirique considérable (737 langues de 14 familles), la stabilité relative des mots de la liste de Swadesh et a proposé un critère pour la quantifier de manière absolue. Cela a permis de poser les bases d'une comparaison de la stabilité diachronique des mots dans les langues du monde, et a créé un nouveau champ pour la sémantique diachronique.

La liste de stabilité maximale obtenue par Serge Starostin est présentée dans Tableau 7)<sup>10</sup>.

Cette liste, fondée sur des moyennes, pose sans aucun doute des problèmes. La notion «boire» figure parmi les dix premières dans les langues khoïsan et les langues austronésiennes, mais dans les langues altaïques il occupe la centième et dernière place. Le mot signifiant «nom» est, dans les langues thaï non pas dans la première, mais dans la dernière dizaine (tout comme dans les langues atlantiques, d'ailleurs). Dans l'ensemble, pour chacune des familles de langues, la place des différentes significations dans telle ou telle dizaine diffère à 75% (!) de celle

<sup>8</sup> Martin Haspelmath, 2008, Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability, in Th. Tolz, D. Bakker & R. Salas Palomo (eds.), *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes*, Berlin: Mouton de Gruyter, p. 43-62, cité d'après [https://www.academia.edu/3015851/Loanword\\_typology\\_Steps\\_toward\\_a\\_systematic\\_cross-linguistic\\_study\\_of\\_lexical\\_borrowability](https://www.academia.edu/3015851/Loanword_typology_Steps_toward_a_systematic_cross-linguistic_study_of_lexical_borrowability)

<sup>9</sup> Starostin, S. A., 2007. Определение устойчивости базисной лексики. [Estimation de la stabilité du lexique de base], *Travaux de linguistique*, Moscow, YSK, p. 827-839.

<sup>10</sup> Dans le projet global «La tour de Babel» créé en 1988 par Serge Starostin et qui connaît aujourd'hui un développement actif, les 50 mots les plus stables sont utilisés avec 9 modifications : les notions avec (-) de la première moitié de la liste sont exclues et remplacées par les notions en gras avec (+) <http://newstar.rinet.ru/babel.php?lan=en>.

qu'elles ont dans le tableau des moyennes. Cependant, les résultats concernant différentes familles sont pour beaucoup convergents. Ainsi, toutes les familles étudiées par Serge Starostin, sans exception, montrent une stabilité beaucoup plus grande dans la première moitié de la liste que dans la deuxième. Il est intéressant de constater que les deux extrémités de la liste ont la valeur la plus universelle. Les notions les plus stables et les moins stables de la liste universelle gardent cette caractéristique dans la majorité des familles de langues. La partie médiane de la liste est la moins universelle. Au regard de l'index de stabilité, les langues de l'Eurasie divergent systématiquement des langues d'Afrique<sup>11</sup>. De plus, il ne faut pas oublier que le lexique le plus stable contient souvent des mots qui ne figurent pas dans la liste de Swadesh.

Tableau 7

	1	2	3	4	5	6	7	8	9	10
1-10	we	two	I	eye	you sg	who	fire	tongue	stone	name
11-20	hand	what	die	heart	drink	dog	louse	moon	nail	blood
21-30	one	tooth	new	dry	( <i>liver-</i> )	eat	tail	( <i>this -</i> )	hair	water
31-40	nose	no	mouth	( <i>full -</i> )	ear	(that)	bird	bone	sun	smoke
41-50	( <i>stand-</i> )	tree	ashes	( <i>give -</i> )	rain	star	( <i>fish -</i> )	( <i>neck -</i> )	( <i>breast-</i> )	leaf
51-60	come	<i>kill +</i>	<i>foot +</i>	sit	root	<i>horn +</i>	fly	<i>hear +</i>	skin	long
61-70	road	know	say	<i>egg +</i>	seed	knee	<i>black +</i>	<i>head +</i>	sleep	burn
71-80	ground	feather	swim	white	bite	<i>meat +</i>	fat	man	person	all
81-90	<i>night +</i>	see	go	hot	red	cold	woman	round	lie	cloud
91-100	big	bark	sand	good	many	yellow	mountain	green	belly	little

Examinons un cas typique avec la famille des langues atlantiques. Ces langues comportent 10 groupes principaux. Présentons les racines atlantiques les plus stables selon notre base étymologique<sup>12</sup>, en supposant que plus le nombre de groupes dans lesquels la racine représentée est important, et plus la racine est stable. Limitons-nous aux significations entrant dans au moins 7 groupes. Dans notre base de données les racines en question sont en tout et pour tout au nombre de 64 (les langues atlantiques sont extrêmement divergentes). La distribution des significations par groupe sont les suivantes :

<sup>11</sup> L'analyse détaillée des données obtenues par Starostin figure dans Konstantin Pozdniakov, 2014, [Discussion] On the threshold of relationship and the "stability index" of basic lexicon in mass comparison: Atlantic languages, *Journal of Language Relationship* 11, p. 187-237 (en russe).

<sup>12</sup> La base de données étymologique atlantique Pozdniakov-Segerer comprend 3550 séries étymologiques de différente profondeur chronologique (février 2016).

10 groupes : **Swadesh** : œil, tête.

9 groupes : **Swadesh** : soleil/jour, manger, oreille, pluie, écorce 1 ; **Hors de Swadesh** : banane, riz.

8 groupes : **Swadesh** : cheveu, langue, étoile, pleuvoir/pluie, **Hors de Swadesh** : nuque, visage/front, attacher/nouer, coudre, éplucher, éternuer, crocodile, jeune fille, lancer, piège/tendre, piler, piquer, prendre.

7 groupes : **Swadesh** : ISG, arbre/médicament, foie, viande, brûler, écorce 2, nager, **Hors de Swadesh** : accoucher, aigu/piquer, année, araignée/tisserand, aveugle, baobab, ceinture à grimper, chanter, chèvre, coude, couvrir, crachat, déchirer, devant, échanger, enfler/abcès, étendre, grand-parent, sorcier, suspendre, monter, planter, riz cuit, semer/planter, souffler, source, sourire, tailler, tas, voir.

Ainsi, notre estimation montre que parmi les 64 racines les plus stables seules 13 figurent dans la liste de Swadesh.

Si j'apporte ici ces nouvelles données, ce n'est pas en vue de soumettre la liste de Swadesh à une ènième révision et proposer aux indoeuropéanistes d'inclure dans la liste les notions «baobab», «crocodile» ou «ceinture à grimper». Le rêve d'une liste universelle adaptée à toutes les langues du monde est illusoire. L'avantage principal de la liste hiérarchisée de Swadesh réside dans le fait qu'elle est approuvée pour des centaines de langues, ce qui signifie que les résultats obtenus par les chercheurs sont compatibles. De plus, l'élaboration fine des listes de Swadesh, dont s'occupe brillamment depuis de nombreuses années le projet global Starling n'exclut pas des recherches typologiques sur l'ensemble du lexique, y compris le lexique culturel.

Arrêtons-nous sur un point important. Dans la liste ci-dessus concernant les langues atlantiques on trouve deux racines pour le sens «écorce» (écorce 1 et écorce 2), une racine «planter» distincte d'une racine «semmer/planter», et de nombreuses notions (ou plutôt la majorité d'entre elles) ont des contours très vagues comme par exemple «araignée/tisserand». Sachant que pour les langues modernes la sélection des mots à inclure dans les listes est déjà une tâche complexe, on comprend combien elle est ardue lorsqu'il s'agit de reconstruction «step-by-step» où le sens des mots dans la protolangue se déduit de notions reconstruites pour chacun des groupes de la famille. C'est ce problème très complexe qu'aborde un des articles de ce volume, écrit par un des leaders de l'école russe d'étude de la parenté linguistique éloignée (macro-comparatisme), Georgui Starostin.

##### 5. CLASSIFICATION GÉNÉALOGIQUE ET MODÈLES DE RAMIFICATION DES LANGUES

Lorsque nous disons que les langues divergent entre elles, il faut se rappeler qu'une langue diverge avant tout d'elle-même en se transformant. C'est pour cela que la question récurrente de savoir depuis quand existe une langue donnée, par exemple le français ou le swahili, appelle vraisemblablement toujours deux réponses qui ne sont pas contradictoires, quelle que soit la langue concernée : 1) des dizaines des milliers d'années, c'est-à-dire depuis que la Langue existe, puisqu'elle en constitue le prolongement, et 2) environ 500 ans, sachant que c'est le temps qu'il faut pour qu'une langue accumule un nombre critique de

divergences par rapport à son état précédent. La question éternelle de la distinction entre langue et dialecte est une question technique pour un comparatiste. En outrant le propos on peut reprendre des observations souvent énoncées : si les idiomes passent sous la barre des 90-95% de correspondances dans le lexique de base de Swadesh, les différents dialectes deviennent différentes langues, et environ tous les 100 ou 200 ans un mot Swadesh disparaît de la langue. Remarquons que ce fait est à la base d'un malentendu de plus, qui est souvent répété dans la littérature : si on affirme qu'en 1000 ans 10% du lexique de base se perd, il semble évident que la limite de portée du comparatisme doit être de 10000 ans, au moment où 100% des mots ont été perdus. Cela n'est pas vrai, pour au moins deux raisons.

Premièrement, comme nous l'avons dit, les mots de la liste du lexique de base se caractérisent par différents degrés de stabilité. Les moins stables disparaissent en premier, et les plus stables peuvent, semble-t-il, ne pas disparaître pendant des dizaines de milliers d'années.

Deuxièmement, il est facile de franchir cette «barrière» grâce à la procédure de reconstruction «step-by-step». Toute langue indo-européenne contemporaine (à l'exception, peut-être de l'albanais) a conservé environ 50% du lexique de base de la protolangue (d'où il suit que les correspondances lexicales entre deux familles de branches différentes est de 25 à 30%). Cependant, grâce au fait que les langues apparentées sont généralement nombreuses, la liste du lexique indo-européen de base est reconstruit de manière fiable à 100%, ce qui signifie que nous n'avons rien perdu en plus de 6000 ans depuis la fin du proto-indo-européen et que nous pouvons éventuellement remonter plus loin dans le temps, si on le souhaite.

Ainsi, la langue a besoin de temps pour accumuler des différences systématiques par rapport à la protolangue et parvenir aux différences régulières avec les autres idiomes qui ont divergé d'elle. La méthode historique et comparée ne fonctionne pratiquement pas pour l'étude des dialectes (une des raisons est l'absence dans la majorité des cas de proto-dialecte, dont les autres dialectes divergeraient). Mais en mille ans de développement indépendant les différences deviennent régulières et systématiques, et le meilleur moyen pour décrire la formation de ces distinctions est toujours le modèle du graphe simple ou de l'«arbre généalogique».

Un autre problème réside dans le fait que la formation réellement indépendante de langues-filles se produit en réalité assez rarement. Il y a certes des cas de type «laboratoire» dans lesquels on peut supposer qu'il existe non pas seulement un siècle, mais même une année, une heure ou une minute à laquelle deux groupes humains qui parlaient une même langue ont commencé un cheminement totalement indépendant de changements linguistiques. Ce moment que nous ne pouvons pas calculer a probablement eu lieu lorsque la flottille de navires antiques d'Asie du Sud-Est ont levé l'ancre afin d'atteindre en quelques mois Madagascar, sur les rives de l'Afrique, et ne plus jamais retourner dans leur patrie historique. C'est aussi ce qui s'est produit pour le rapanui lorsqu'un groupe de Polynésiens a quitté il y a environ mille ans les Îles Marquises pour atteindre l'Île de Pâques et y rester pour toujours. Ce type de scénario est assez

exceptionnel. Il arrive souvent que des ethnies connaissent des changements d'influences culturelles et linguistiques sans avoir à se déplacer, et s'ils le font, comme par exemples les nomades peuls qui ont quitté avec leurs troupeaux leur territoire et leurs voisins pour y revenir quelques siècles plus tard, ils retrouvent un paysage cultural totalement transformé. Dans ce cas il est extrêmement difficile, ou même parfois impossible, de distinguer clairement entre les réflexes réguliers des proto-formes et les emprunts anciens (a fortiori s'ils proviennent de langues proches). Cela a pour conséquence que le modèle de l'arbre, qui ne fait que refléter nos connaissances de l'histoire des langues, manifeste des contradictions flagrantes, et que le comparatisme voudra remplacer ce modèle par un autre, plus adéquat à la situation.

Le modèle qui se différencie le plus de celui de l'arbre est le modèle classique des ondes de Johannes Schmidt, qui décrit sans aucun doute beaucoup mieux la variation dialectale des langues et leurs mécanismes de contact. On connaît plusieurs tentatives de concilier ces deux modèles en créant un qui serait à même de traiter efficacement toutes les situations. Dans ce recueil une nouvelle tentative dans ce sens est présentée par Alexandre François.

Alexandre François n'est pas le seul auteur de ce volume à aborder le problème de la classification. Aucun des articles du volume n'est directement et entièrement consacré à cette question très complexe, mais la question de la classification est abordée dans les articles de G. Jacques, C. Boyeldieu, C. Rilly, J.-P. Demoule, G. Starostin, K. Pozdniakov, G. Segerer, V. Vydrin. De fait, les problèmes de classification généalogique sont loin de se limiter à la question du choix du modèle de représentation le plus adéquat. La question essentielle est de savoir sur quel principe la construire. Nous remarquons une fois de plus que les procédures de la méthode historique et comparée ne donnent pas de réponse à cette question essentielle. A cet égard la lexicostatistique n'est qu'une des approches envisageables, à côté d'autres. Plus encore, parfois une isoglosse grammaticale a plus de valeur pour la classification que deux ou trois racines partagées du vocabulaire de base, comme c'est le cas dans la branche des langues atlantiques Nord pour lesquelles la meilleure isoglosse est la présence d'un système d'alternances consonantiques lié aux classes nominales, inexistant dans l'autre branche atlantique (Bak). J'évoquerai ici une expérience personnelle. Pour rédiger l'article sur la classification des langues atlantiques<sup>13</sup> nous nous sommes tout d'abord adressé aux spécialistes de lexicostatistique (Georgui Starostin) qui nous ont apporté leur aide. Sur la base des matrices que nous avions renseignées, nous avons pu aboutir à un arbre mais nous avons décidé de ne pas l'inclure dans l'article, pour la raison suivante. La classification lexicostatistique a fidèlement reflété les regroupements les plus évidents et déjà bien connus (nul besoin d'être spécialiste pour conclure que le français est plus proche de l'espagnol que du russe). Pour ce qui est des regroupements de deuxième ordre, la lexicostatistique nous a proposé des variantes qui sur beaucoup de

---

<sup>13</sup> Pozdniakov K. & Segerer G. 2016, à paraître, A new classification of Atlantic languages, in F. Luepke (ed), *The [Oxford] Guide to the Atlantic Languages of West Africa*, Oxford University Press.



points contredisaient les autres données dont nous disposions. Je souligne encore une fois qu'il ne s'agit pas ici de lacunes de l'approche de lexicostatistique, mais avant tout du niveau de nos connaissances dans le domaine des études atlantiques : 1) la reconstruction «step-by-step» du lexique atlantique n'est pas encore aboutie, et les arbitrages concernant la parenté des formes ont dû être faits «à vue d'œil» ; 2) il s'agit d'une famille particulièrement divergente, pour laquelle la glottochronologie propose l'âge de 9000 ans. Cela signifie que les conclusions sur les premières ramifications de la famille dépendent finalement de l'interprétation de deux ou trois racines. Au final nous avons préféré nous appuyer sur les innovations de différents types, lexicales, phonétiques, morphologiques et sémantiques. Ce principe a été formulé au 19<sup>ème</sup> siècle par Leskien et est cité dans ce volume par Alexandre François : «*to distinguish shared retentions from shared innovations, as only the latter can tell us anything about the genealogical history of a linguistic family*». A mon sens c'est justement cette distinction essentielle entre des retentions et des innovations qui est passée sous silence avec l'approche de lexicostatistique.

L'article de Pascal Boyeldieu, qui applique aux langues Sara-Bongi-Bagirmi un travail réalisé avec Pierre Nougayrol, présente une autre approche de la classification très intéressante. Contrairement à la lexicostatistique qui, comme nous l'avons vu, écarte le lexique culturel de l'analyse, Boyeldieu-Nougayrol le placent au centre de la classification généalogique. Le principe est de répartir le lexique culturel en groupes thématiques pour suivre la répartition des isoglosses pour chacun de ces groupes, avant de procéder à une étude comparative des résultats obtenus, ce qui lui permet de conclure «sur le caractère fortement renouvelable du vocabulaire lié à un environnement spécifique».

#### 6. OÙ ET PAR QUI LA PROTOLANGUE A-T-ELLE ÉTÉ PARLÉE ?

Soulignons une fois de plus que la méthode historique et comparée n'est liée à aucune de ces deux questions. C'est pourquoi la célèbre monographie en deux volumes de Gamkrelidze et Ivanov «La langue indo-européenne et les Indo-Européens»<sup>14</sup>, du fait de son titre provocateur, a suscité une discussion animée bien avant sa publication effective. Le point de vue traditionnel sur cette question consiste à penser que nous pouvons mettre en évidence la parenté des langues et même reconstruire une protolangue, mais que nous n'avons aucun instrument fiable pour dire quel(s) peuple(s) a/ont parlé(s) cette langue et sur quel territoire. Pour répondre à ces questions les efforts conjugués de linguistes, archéologues, paléobotanistes, paléozoologues, généticiens, anthropologues et bien d'autres spécialistes concernés par l'histoire ancienne.

A cet égard, parmi les publications récentes il faut mentionner la monographie polémique de Jean-Paul Demoule<sup>15</sup>, où l'éminent historien et archéologue évoque

<sup>14</sup> T. V. Gamkrelidze & V. V. Ivanov, 1984, *Индоевропейский язык и индоевропейцы* [La langue indo-européenne et les Indo-Européens], 2 vol, Tbilisi, Editions Université de Tbilissi.

<sup>15</sup> Demoule J.-P., 2014, *Mais où sont passés les Indo-Européens ?*, Paris, Ed. du Seuil, surtout le chapitre 11 «Ce que nous dit l'archéologie aujourd'hui» et la suite.

plusieurs problèmes de ce domaine interdisciplinaire. J.-P. Demoule n'a malheureusement pas trouvé dans ce volume l'espace pour développer son expertise d'archéologue, concentrant son article sur la problématique purement linguistique. Pour le linguiste, son article présente notamment l'intérêt de livrer une synthèse quasi-idéale de la critique traditionnellement adressée à la méthode historique et comparée, y compris vis-à-vis de points de vue que je tente de défendre dans cette présentation.

Ce volume comporte également une publication qui illustre le vecteur inverse, montrant qu'un linguiste peut se permettre d'aborder en connaissance de cause les données de l'archéologie, nommément Claude Rilly. En s'appuyant sur les données interdisciplinaires, l'auteur place le proto-soudanique Est (PES) au Sud de l'Égypte et propose les voies concrètes de la dispersion des langues de cette famille.

Pour conclure, rappelons encore que le comparatiste ne dispose pas aujourd'hui d'outils fiables pour dire qui parlait la protolangue et dans quel territoire, même si certaines observations, notamment à caractère typologique, ont été avancées. Par exemple, en termes typologiques, il peut être fondé de rechercher la patrie originelle d'une famille linguistique étendue et typologiquement divergente dans un lieu où on peut trouver aujourd'hui une grande variation linguistique. Ainsi, la patrie originelle de la famille Niger-Congo, le plus vaste groupement génétique de langues connu à ce jour, se trouvait vraisemblablement dans le delta du fleuve Niger au Nigeria et au Cameroun voisin, territoires dans lesquels on trouve une concentration exceptionnelle de langues de différentes branches Niger-Congo.

## 7. BASES DE DONNÉES ET MÉTHODES POUR LA RECHERCHE COMPARATISTE

Dans le domaine des bases de données comparatistes, il est clair que se présentent à nous des possibilités de recherche totalement nouvelles, qui n'étaient pas accessibles à nos prédécesseurs, et qui commencent déjà à porter leurs fruits. Dès que ces multiples bases impressionnantes deviendront compatibles entre elles (ce que n'est pas encore le cas aujourd'hui) nous aurons une nouvelle révolution informatique dans la discipline du comparatisme.

Cependant, certaines bases déjà disponibles aujourd'hui sont tellement volumineuses qu'elles peuvent présenter autant de problèmes que d'avantages pour les chercheurs. Hier on savait très peu, aujourd'hui on sait «trop». Pour le comparatiste qui veut aboutir à un résultat, cela multiplie considérablement les écueils, pièges, contre-arguments contre toute hypothèse de bon sens en comparaison avec la situation d'il y a encore vingt ans. Par expérience, je sais que cette avalanche de données primaires décourage les chercheurs les plus audacieux, les mettant au pied d'un dilemme : digérer toute cette avalanche avant toute publication, ou fermer les yeux sur la masse des contradictions aimablement mises à notre disposition par ces énormes bases de données. Et s'il faut sacrifier certaines contradictions, alors lesquelles ?

Sur ces questions l'article de G. Segerer où l'auteur présente sa base intitulée RefLex mérite l'attention. Il existe très peu de bases construites non pas par des

informaticiens, mais par les usagers eux-mêmes, des comparatistes qui savent de première main ce qu'il faut inclure dans la base et comment il faut organiser l'information accumulée. La meilleure base (et aussi logiciel de recherche) de ce genre est toujours à ma connaissance StarLing, la remarquable base construite par Sergey Starostin il y a environ 40 ans<sup>16</sup>. Il est inutile de comparer StarLing et RefLex – ce sont des outils très différents, car orientés vers des objectifs différents. Notons seulement que la base RefLex étant la source lexicographique la plus importante pour les langues d'Afrique (elle accumule plus d'un million des mots avec accès direct à partir de chaque mot à la source en PDF) offre au comparatiste des conditions de travail d'un confort inimaginable il y a encore dix ans.

Le dernier article du volume (Konstantin Pozdniakov) fait le bilan du travail expérimental conduit par l'auteur depuis de nombreuses années dans le domaine de l'application de la statistique élémentaire à l'étude des questions du comparatisme. Les éléments du «comparatisme quantitatif» que j'y propose s'appuient sur un postulat simple : chaque changement dans la langue provoque automatiquement la modification d'une de ses propriétés statistiques, ce dont il suit que chaque variation «quantitative» des langues apparentées doit trouver un explication «qualificative» dans la diachronie des langues.

Au total, les articles proposés au lecteur de ce volume englobent une grande partie de la problématique historique et comparée et reflètent dans une grande mesure les tendances actuelles en comparatisme et reconstruction. Les familles de langues représentées dans le volume sont :

Langues Vanuatu, famille austronésienne (Alexandre François),  
 Famille sino-tibétaine (Guillaume Jacques),  
 Famille indo-européenne (Romain Garnier, Petr Kocharov, Jean-Paul Demoule),  
 Langues nakh, famille des langues caucasiennes (George Starostin),  
 Langues tchadiques, famille afro-asiatique (Victor Porkchomosky),  
 Langues soudaniques Est, macro-famille Nilo-Saharienne (Claude Rilly),  
 Langues Sara-Bongo-Bagirmi, macro-famille Nilo-Saharienne (Pascal Boyeldieu),  
 Langues mande, macro-famille Niger-Congo (Valentin Vydrin),  
 Langues atlantiques, macro-famille Niger-Congo (Guillaume Segerer, Konstantin Pozdniakov).

Pour conclure, je voudrais une fois de plus insister sur le fait que grâce à la méthode comparée et historique, les comparatistes disposent d'un outil puissant pour prouver la parenté des morphèmes et des langues, ce qui est déjà beaucoup. En allant plus loin, on risque de 1) reprocher à la méthode classique ce qu'elle n'est pas destinée à faire, et 2) s'abriter derrière sa réputation sans failles pour donner du poids à des idées qui ne lui sont en rien liées<sup>17</sup>.

<sup>16</sup> Voir la présentation de cette base sur <http://starling.rinet.ru/>.

<sup>17</sup> Un exemple de ce genre très connu parmi les africanistes est le travail de Ch. Ehret (par exemple : Christopher Ehret. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln, Rüdiger Köppe Verlag, 2001) que Gérard Philippon a discuté dans notre séminaire.

Nous remercions tous les évaluateurs qui ont contribué à la qualité de ce volume. Rappelons que chacun des articles est la version développée d'une présentation dans notre séminaire, et nous espérons que l'esprit de discussion active que nous avons cherché à impulser dans ce séminaire transparait dans ce volume.