

INTRODUCTION

Presque toutes les langues européennes et un grand nombre de langues d'Asie Mineure et d'Inde appartiennent à une même famille linguistique : la famille indo-européenne.

Cette constatation, toute récente puisqu'elle a été faite il y a moins de deux siècles, coïncide avec la naissance de la linguistique moderne. C'est en effet la découverte par les occidentaux du sanskrit, langue sacrée de l'Inde, dont les premiers témoignages sont antérieurs au IXe siècle av. J.-C., qui a permis de mettre en évidence à la fois l'existence et l'extension géographique de la famille indo-européenne.

On peut citer, parmi les langues actuellement parlées, en allant de l'occident vers l'orient, les langues romanes, celtiques, anglo-saxonnes, scandinaves, germaniques, l'albanais, le grec, le lithuanien, le lette, les langues slaves, l'arménien, le kurde, le persan, l'afghan et la centaine de langues d'origine indo-européenne de la péninsule indienne : hindi, hindoustani, marathe, pendjabi, goujrati, rajasthani, népalais, singhalais, etc. Parmi les langues aujourd'hui disparues ou qui se continuent dans les langues vivantes actuelles, mentionnons le gaulois, les langues italiques (latin, vénète, osque, ombrien, etc.), le gotique, le vieux prussien, le thrace, le phrygien, les langues anatoliennes (hittite, louvite, palaïte, lydien, lycien), le vieux perse, l'avestique, le sanskrit et le tokharien.

I. Historique des études indo-européennes.

Les premières observations concernant des similitudes entre le sanskrit et une langue européenne ont été faites par l'italien Sasseti qui, dans les lettres qu'il écrivait de Goa entre 1583 et 1588, avait noté un certain nombre de correspondances entre le sanskrit et sa langue maternelle : sk.

sarpa, it. *serpe* ; sk. *sds*, it. *sei* ; sk. *saptá*, it. *sette* ; sk. *astau*, it. *otto* ; sk. *náva*, it. *nove*, etc.

En 1767, un jésuite français, Coeurdoux, installé à Pondichéry, soulignait, dans une note à l'Académie des inscriptions, les "curieuses analogies existant entre la langue sanscrouane et le latin et le grec". Il citait les exemples suivants : sk. *dāna*, lat. *donum*, sk. *vidhāvā*, lat. *vidua*, sk. *āsmi*, gr. *είμι*. Cette note qui posait le problème des correspondances mais n'apportait pas d'explication, n'eut guère d'écho à l'époque.

La solution est entrevue, quoique non encore démontrée en 1789, par un magistrat anglais, sir William Jones, dans un discours prononcé devant la *Royal Asiatic society*, à Calcutta. Reconnaisant que le sanskrit a, avec le latin et le grec, "tant dans les racines des mots que dans les formes grammaticales, une affinité trop forte pour qu'elle puisse être le produit d'un hasard", il ajoutait : "aucun philologue ne pourrait examiner ces langues sans acquérir la conviction qu'elles sont en fait issues d'une source commune, laquelle, peut-être, n'existe plus. Il y a du reste une raison similaire, quoique pas tout à fait aussi contraignante, pour supposer que le gothique et le celtique, s'ils ont été mêlés par la suite avec un parler différent, n'en descendent pas moins en définitive de la même origine que le sanskrit ; on pourrait ajouter en outre à cette famille le vieux perse".

Mais, en l'absence d'une méthode scientifique permettant de pousser plus loin la comparaison, cette bonne intuition demeura, comme la note précédente, sans suite immédiate.

Il faut attendre la création, en 1795, de l'Ecole nationale des langues orientales vivantes pour que le sanskrit soit enseigné en Occident, à Paris. De toute l'Europe, et en particulier d'Allemagne, des linguistes possédant une solide formation classique (latin et grec) et connaissant des langues germaniques et slaves, vont venir s'initier au sanskrit et constater que l'hypothèse d'une origine commune est, de loin, plus logique que le recours à d'hypothétiques emprunts. C'est à l'Ecole nationale des langues orientales que Frédéric Schlegel se forma au sanskrit avant d'écrire son fameux livre *Ueber die Sprache und Weisheit der Inder* (1808). Il y déclare que la comparaison des langues "nous donnera des informations entièrement nouvelles sur la généalogie du langage,

tout comme l'anatomie comparée a jeté de la lumière sur l'histoire naturelle".

Le terrain, on le voit, est désormais préparé, et ce n'est pas un hasard si la démonstration est apportée peu après, quasi simultanément et tout à fait indépendamment, par un Danois, Rasmus Rask (1814) et par un Allemand, Franz Bopp (1816). Cependant, malgré la rigueur de la méthode du premier, son ouvrage, publié en danois, en 1818 seulement, ne connaît pas la diffusion de celui de Bopp. C'est finalement Franz Bopp qui restera le fondateur de la grammaire comparée des langues indo-européennes, avec son mémoire *Ueber das conjugationssystem der sanskritsprache, in vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache*. Il donnera par la suite une plus grande extension à son étude avec les deux éditions de sa *Vergleichende Grammatik* (1833-1849 et 1857). Il intégrera, outre le sanskrit, le grec, le latin, le persan et le germanique, successivement le "zend", le lituanien, le vieux-slave et enfin l'arménien. Il échouera, en revanche, dans sa tentative d'intégrer aussi le malayo-polynésien et le géorgien. L'inventaire des langues indo-européennes vivantes était ainsi, dès cette époque, pratiquement achevé. Il n'en était pas de même des langues anciennes.

En effet, au début du XXe siècle, la découverte, au Turkestan chinois, d'un grand nombre de textes d'une langue aujourd'hui disparue, le tokharien (avec deux variétés, dites tokharien A et B), puis, à partir de 1916, de plusieurs langues répandues en Asie Mineure au II^e millénaire avant notre ère, hittite, louvite, lycien, lydien, allait étendre encore la famille indo-européenne, permettre d'améliorer les méthodes de la grammaire comparée et même de vérifier certaines hypothèses. Ainsi, une hypothèse de F. de Saussure postulant l'existence, en indo-européen, d'un phonème appelé *chva* et transcrit par un e renversé (*ð*), n'a été confirmée qu'après la mort du linguiste genevois, lors du déchiffrement du hittite : celui-ci conserve en effet, nettement transcrits par les caractères cunéiformes, les phonèmes que, pour des raisons purement théoriques, F. de Saussure avait été amené à postuler.

Sur les matériaux ainsi accumulés depuis plus d'un siècle et demi, ont travaillé des savants remarquables parmi lesquels il faut citer J. Grimm qui a donné son nom à une loi

phonétique fondamentale, A. Schleicher, K. Brugmann, A. Meillet, J. Vendryes, E. Benveniste.

II. La méthode de la grammaire comparée.

La grammaire comparée repose donc sur l'observation d'un grand nombre de concordances entre des langues que tout semble apparemment séparer : la localisation géographique qui s'étend sur plus de 10 000 km dans l'ancien monde, comme les époques auxquelles elles ont été parlées (du III^e millénaire avant J.-C., à nos jours). Ces concordances touchent tous les domaines de la langue (phonétique, morphologie, syntaxe, vocabulaire) et amènent à postuler l'existence, à une époque préhistorique, d'une langue qui aurait servi de point de départ à chacune d'entre elles.

Cependant, le terme de "grammaire comparée" ne doit pas abuser : il ne s'agit pas de reconstruire la grammaire de l'indo-européen. Comme l'a affirmé Antoine Meillet, "l'indo-européen est inconnu, et les concordances sont les seules réalités qu'ait à étudier le comparatiste. La grammaire comparée n'a pas pour but de reconstruire l'indo-européen, mais, grâce à la détermination des éléments communs indiqués par les concordances, de mettre en évidence ce qui, dans chacun de ces idiomes historiquement attestés, est la continuation d'une forme ancienne de la langue, et ce qui est dû à un développement propre et original".

Il est bien évident que, dans sa remontée du temps vers la langue d'origine, le comparatiste prend appui sur les langues les plus anciennement attestées. Ainsi, puisque l'on sait que le portugais, l'espagnol, le catalan, le français, le provençal, l'italien, le roumain, etc., sont des différenciations d'une même langue, le latin, l'indo-européaniste ne prendra en compte que ce dernier. De même, il négligera l'anglais, le hollandais, le danois ou l'allemand moderne, pour ne porter son attention que sur le gothique (IV^e siècle ap. J.-C.), les inscriptions runiques (II^e siècle ap. J.-C.), le haut et le bas allemand (IX^e siècle), le vieil anglais (IX^e siècle), etc.

La méthode fondamentale qui a permis la reconstruction d'un grand nombre de racines indo-européennes découle de l'observation de la régularité des lois phonétiques. Ainsi,

la loi de Grimm établit que les consonnes occlusives latines, p, t, k - et grecques π, τ, α - correspondent aux fricatives germaniques f, þ, h, cependant que les consonnes germaniques p, t, k, sont représentées en latin par les consonnes sonores b, d, g (gr. β, δ, γ). Par rapport à l'indo-européen, représenté ici par le latin et le grec, le système des consonnes du germanique s'est donc déplacé d'un degré.

On obtient ainsi les correspondances suivantes pour le nom du père : sk. *pitár-*, grec *πατήρ*, lat. *pater*, mais got. *fadar*; et le nom du pied se modifie pareillement : accusatif sk. *pādam*, grec *πόδα*, lat. *pedem*, mais got. *fotu*.

La régularité est en principe parfaite, c'est-à-dire que, placés dans des conditions semblables, tous les sons identiques d'une langue évoluent en même temps et de la même manière. Cependant, des influences analogiques diverses peuvent venir troubler la constance des lois phonétiques, surtout pour des langues qui ont été séparées pendant plusieurs millénaires.

A cet égard, la grammaire comparée des langues romanes, dont la séparation remonte à moins de deux millénaires, à partir d'une langue - le latin - qui est fort bien connue, reste l'idéal impossible à atteindre pour un indo-européaniste. Dans le cadre de ce groupe de langues, il est en effet possible d'opérer une reconstruction par la comparaison et, *en même temps*, de vérifier, en recourant au latin, la justesse de la méthode. On constate alors, dans ces conditions privilégiées, que la méthode ne permet pas, dans tous les cas, de restituer la forme de départ. Ainsi, les correspondances pourtant fort régulières suivantes :

fr. <i>huit</i> ,	it. <i>otto</i> ,	esp. <i>ocho</i> ,	roum. <i>opt</i> ,
fr. <i>fait</i> ,	it. <i>fatto</i>	esp. <i>hecho</i> ,	roum <i>fapt</i> ,

permettent certes de trouver qu'au français *nuit* va correspondre l'italien *notte*, l'espagnol *noche*, le roumain *no(a)pte*, ou encore qu'au français *lait* va correspondre l'italien *latte*, l'espagnol *leche*, le roumain *lapt*, etc. Mais à aucun moment le groupe latin *-ct-* présent dans *octo*, *noctem*, *factum*, **lactem**, n'apparaît et ne pourra donc être restitué. On comprend ainsi que les comparatistes aient voulu limiter leurs certitudes aux correspondances et se soient refusés à reconstruire véritablement l'indo-européen. L'exemple de Au-

gust Schleicher écrivant, au milieu du siècle passé, dans l'enthousiasme des premières découvertes, une fable en "indo-européen commun" (intitulée *avis akvasas ka* 'la brebis et les chevaux') est une exception qui révèle l'impossible idéal auquel aspiraient les premiers comparatistes.

Peut-on espérer un jour découvrir des textes écrits en indo-européen, comme on a retrouvé des tablettes en langue hittite ou louvite ? La chose est peu vraisemblable. En effet, les langues anatoliennes (hittite, louvite, lycien, lydien) n'ont pu être conservées que parce que des peuples d'origine indo-européenne venus du nord sont entrés en contact avec les civilisations mésopotamienne et élamite au sud, qui possédaient l'écriture. Celle-ci n'apparaissant, sous sa forme pictographique, que vers 3300 av. J.-C. et ne s'étant répandue au Proche-Orient qu'à partir de 2500 av. J.-C. (écriture cunéiforme), c'est-à-dire après la séparation des indo-européens, il est pratiquement exclu - à moins de bouleversement complet de nos connaissances sur ce point - que des découvertes ultérieures nous mettent en présence de traces écrites d'indo-européen commun. Les recherches archéologiques pourraient cependant nous apporter des écrits antérieurs de quelques siècles à ceux que nous possédons déjà et qui remontent à l'ancien empire hittite (1650-1500 av. J.-C.). Elles pourraient surtout nous livrer d'autres témoignages matériels de l'existence de populations indo-européennes.

En l'absence de renseignements directs, les correspondances linguistiques observées, surtout au niveau de la morphologie et du vocabulaire, permettent malgré tout de se faire une idée assez précise, non seulement de la langue, mais aussi des conceptions idéologiques, politiques, sociales, économiques des populations indo-européennes.

III. Les caractéristiques de l'indo-européen.

Le système phonétique de l'indo-européen différait sensiblement de celui que connaissent les langues occidentales modernes. Il comprenait certes des consonnes (sourdes, sonores, aspirées) et un jeu d'alternances vocaliques de trois voyelles (e, o, a, brèves ou longues, auxquelles il faut ajouter le chva *ǝ*), mais aussi un nombre élevé de sonantes qui

pouvaient jouer le rôle de voyelles ou de consonnes : *y, w, r, l, m, n*, et les différents types de laryngales.

La morphologie est d'une grande complexité. Elle se répartit en deux grandes catégories : le nom et le verbe, comme les langues indo-européennes actuelles. Mais la première comporte une déclinaison à huit cas : nominatif (et ergatif pour les animés), vocatif, accusatif, génitif, datif, locatif, instrumental et ablatif). La distinction du masculin et du féminin n'est pas toujours marquée : elle s'est constituée à une date récente. A l'origine, le genre oppose l'animé à l'inanimé. Le nom possède trois nombres : le singulier, le duel et le pluriel. Il en est de même du verbe qui, en outre, comme dans nos langues, distingue trois personnes. La caractéristique principale du système "temporel", c'est qu'il n'exprime pas le temps comme nous l'entendons aujourd'hui (passé, présent, futur), mais le degré d'accomplissement de l'action. En ce sens, il est plus proche de la représentation verbale des langues sémitiques que de celle des langues indo-européennes occidentales.

Le verbe pouvait être mis à la voix active si le sujet était entièrement maître de l'action ou à la voix moyenne dès que cette maîtrise lui échappait, même partiellement. Le latin représente en partie cet état de langue avec deux voix seulement : une voix active (ex : *amo*) et une voix moyenne ne faisant pas de distinction entre les différents degrés de passivité : *amor* (passivité totale) et *sequor* (passivité partielle, déponent) sont tous deux à la voix moyenne, dite aussi médio-passive.

La conjugaison, telle que nous la connaissons, n'existait pas : chaque racine pouvait recevoir des thèmes différents et fort nombreux. C'est par centaines qu'apparaissent les formes possibles d'une même racine si on fléchit tous les thèmes à tous les nombres, à toutes les personnes, à toutes les voix et avec toutes les sortes de désinences. Au cours de l'histoire, une simplification progressive s'est opérée, combinée avec une régularisation du système verbal : seules les langues anciennes de l'Inde, du Moyen-Orient et de Grèce peuvent donner une idée de la complexité de la langue d'origine.

IV. L'expansion de l'indo-européen dans l'espace et dans le temps.

Si la grammaire comparée permet de reconstituer des états de langue anciens, elle n'autorise pas à situer de manière précise la date de la phase indo-européenne commune. L'étude du vocabulaire donne cependant des renseignements précieux : ainsi, le terme **akmen-* formé à partir de la racine **ak-* (= 'aigu') et d'une dérivation en **-men*, désigne à la fois l'enclume et la pierre. On peut penser que l'on est passé de la désignation de la pierre au néolithique à celle de l'enclume au moment du travail du bronze. En revanche, comme les termes qui désignent le fer sont différents dans les langues indo-européennes -donc récents-, alors que ceux qui désignent le cuivre sont communs, la conclusion suivante s'impose : les indo-européens se sont séparés avant l'âge du fer, mais après celui du cuivre.

Les recherches récentes situent l'habitat des indo-européens de la période commune dans le sud de la Russie, près des sites cuprolithiques occupés de manière continue par les civilisations dites de Srednij Stog II (4400-3400 av. J.-C.), d'Usatovo-Mikhaloska (3500-2900 av. J.-C.) et de Jamma (III^e millénaire). Les caractéristiques de ces civilisations découvertes par l'archéologie correspondent bien à celles que la grammaire comparée attribue aux Indo-Européens. Les Aryens seraient partis du site de Jamma au cours du III^e millénaire pour se diriger vers le sud et vers l'ouest. La première branche arrivera en Inde au début du II^e millénaire, cependant qu'une autre partie se fixera en Iran.

Les vagues successives de populations indo-européennes qui coloniseront peu à peu l'Europe entière (Achéens, Hellènes au sud, Celtes, Germains, Baltes à l'ouest et au nord) ne peuvent être reconstituées par la dialectologie des langues indo-européennes occidentales. On ne sait d'où venaient, par exemple, les Latins, cette infime partie des Indo-Européens qui allaient pourtant avoir une destinée exceptionnelle puisqu'après avoir atteint l'océan Atlantique en Gaule et dans la péninsule ibérique, ils allaient le traverser pour découvrir le nouveau monde et y imposer des langues néo-latines, c'est-à-dire, en définitive, quelques-unes des nombreuses variantes de l'"indo-européen actuel". A l'échelle des millénaires, l'expansion des langues romanes, comme celle de l'anglais, en

Amérique et en Afrique, apparaît donc comme le prolongement d'un mouvement continu commencé il y a plus de 5000 ans et qui se déroule encore sous nos yeux.

Alvaro ROCCHETTI